BIAS CORRECTION AND CHANGE MEASUREMENT IN SPATIO-TEMPORAL DATA

A thesis submitted in partial fulfilment of the requirements for the Degree

of Doctor of Philosophy in Statistics

in the University of Canterbury

by Miriam Christine Hodge

University of Canterbury

2012

---

1 Acknowledgments

This thesis would not have been possible without love and unconditional support of my family and especially my father, Brien. His famous refrain, “get the encyclopaedia,” still rings in my head whenever I fact check. My supervisors, Marco and Jennifer, gave me guidance on my research as well as invaluable input on many personal and professional matters. My friends and colleagues listened to my rambling conversations about my thesis and smiling while I talked things out. Tom agreed that New Zealand looked great on the internet and we should go sight unseen. New Zealand provided a sensible and generous student aid programme. Finally, The University of Canterbury admitted me and offered the Doctoral Scholarship that underwrote this work.

Alice

By Shel Silverstein

She drank from a bottle called DRINK ME
And up she grew so tall,
She ate from a plate called TASTE ME
And down she shrank so small.
And so she changed, while other folks
... Never tried nothin at all.

2 Contents

2.1 Table of contents

2.2 Table of figures

2.3 Table of tables

3 Abstract

A simplistic view of a dataset is that it is collection of numbers. In fact data are much more than that and all data are collected at a set place and time. Often either the location, or the time, is fixed within the dataset and one or both are disregarded. When the place and time of the collection are incorporated into the analysis, the result is a spatio-temporal model.

Spatio-temporal data are the focus of this thesis. The majority of the datasets used are radio tracking studies of animals where the objective is to measure the habitat use. Observations are made over a long period of time and a large area. The largest dataset analysed tracks over a hundred animals, in an area larger than 40 square miles, for multiple years. In this context understanding the spatio-temporal relationships between observations is essential. Even data that do not have an obvious spatial component can benefit from spatio-temporal analysis. For example, the data presented on volatility in the stock market do not have an obvious spatial component. The spatial component is the location in the market, not a physical location.

Two different methods for measuring and correcting bias are presented. One method relies on direct modelling of the underlying process being observed. The underlying process is animal movement. A model for animal movement is constructed and used to estimate the missing observations that are thought to be the cause of the bias. The second method does not model the animal movement, but instead relies on a Bayesian Hierarchical Model with some simple assumptions. A long running estimation is used to calculate the most likely result without ever directly estimating the underlying equations.

In the second section of the thesis two methods for measuring change from shifts in both spatial and temporal location are presented. The methods, Large Diffeomorphic Deformation Metric Mapping (LDDMM) and Diffeomorphic Demons (DD), were originally developed for anatomical data and are adapted here for nonparametric regression surfaces. These are the first applications of LDDMM and DD outside of computational anatomy.

4 Abbreviations

AIC. Akaike Information Criterion

AIC_C. Corrected Akaike Information Criterion

AMISE. Asymptotic Mean Integrated Square Error

ATS. Automated Telemetry System

BHM. Bayesian Hierarchical Model

VIX. Chicago Board Options Exchange Market Volatility Index

DD. Diffeomorphic Demons

GIS. Geographic Information System

GPS. Global Positioning System

GARCH. Generalised Autoregressive Conditionally Heteroskedastic

IQR. Inter-Quartile Range

LDDMM. Large Deformation Diffeomorphic Metric Mapping

MCMC. Markov Chain Monte Carlo

MISE. Mean Integrated Square Error

MCP. Minimum Convex Polygon

PCA. Principal Component Analysis

PDE. Partial Differential Equation

PDF. Probability Density Function

REML. Restricted Maximum Likelihood

RSF. Resource Selection Function

SP500. Standard and Poor's 500 index options traded on the Chicago Board of Exchange

TP. Time Periods

5 Introduction

Spatio-temporal statistics incorporates time and place of observations into its analysis. All data have a spatial and a temporal component, but many times, both are not explicitly included in the analysis. If the process being studied evolves over time and varies by location then spatio-temporal analysis can aid in understanding the mechanisms underlying the process. For example, the snap shot of the population of Abu Dhabi taken on the night of 3 October, 2011 (SCAD, 2011; Hui and AlDarmaki, 2012) can be analysed ignoring the temporal component of the data. Likewise population estimates from the quinquennial census of Abu Dhabi begun in 1971 (NBS, 2011) can be analysed ignoring the spatial component. However, spatio-temporal analysis is appropriate to analyse the changes in population density of residence within Abu Dhabi from 1971 to present, because both the spatial and the temporal change are included in the analysis. This thesis explores two topics in spatio-temporal statistics, bias correction and metrics for comparing regression surfaces. It investigates these topics utilising data for radio telemetry tracking of animals, the design of aircraft, ozone measurements, and stock market volatility.

Radio tracking studies of animal populations involve collecting observations that have both a spatial and a temporal context. A radio device is placed on an animal and the animal's location is recorded on a schedule. Each of these observations is called a fix. A fix has a spatial component consisting of latitude longitude and elevation, and a temporal component consisting of the time of the recording. The spatial location of the fix can be cross referenced with a habitat map to determine the type of habitat the animal is using. Non-parametric regression is used as an

analysis tool to describe the habitat used by the animal mapped to the study area in a given year. The result is a series of surfaces, one for each year, which collectively describe the change in habitat use over time. The habitat use maps based on only successful fixes can be biased because the variable of interest, habitat use, can impede the radio signal by which the observation are made. Two methods are presented for bias correction in radio tracking studies.

Two novel methods are presented to compare the surface shapes as a way of describing change. Surfaces are compared using metrics developed for comparing shapes in medical imagery. Metrics based on deformation, a technique developed for anatomical surfaces, are adapted to compare non-parametric regression surfaces. The utility of these methods is explored on several datasets including the design of aircraft, ozone measurements, and stock market volatility.

5.1 Bias Correction for missing fixes in radio telemetry studies

Radio telemetry studies are an example of data well suited for spatio-temporal analysis because both time and location play an integral role in understanding the data; therefore both need consideration during the analysis. Neither the spatial nor the temporal component can be ignored because spatial components such as animal behaviour, radio reception, and habitat are affected by temporal factors. Animals have different behaviours at different times of the day and different seasons of the year. Radio reception is affected by temporal factors. For example, GPS is affected by the relative position of satellites during their daily orbit. Habitat also changes over time due to seasonal changes such as trees dropping leaves. This habitat change can have a spill-on effect on the radio reception because the density of canopy affects radio reception.

Radio telemetry studies have spatial variation in animal behaviour, habitat and radio reception. One popular technique for handling spatial variation during analysis is to chop up the study area

into equally sized grid squares. One early analysis done via this method was locating a lost submarine with a nuclear payload (Cressie and Wikle, 2011). This type of calculation is easily carried out with a computer and underpins much of the analysis done by Geographic Information Systems (GIS). It allows spatial patterns to be easily displayed for each location in the study area.

The location data collected during radio telemetry studies are often combined with habitat maps to estimate habitat use by animals. This leads to biased estimates of habitat use (Rettie and McLoughlin, 1999) because the use of habitat with bad radio reception is underestimated. The two broad categories for bias correction are iterative simulation and weighting. Iterative simulation augments successful fixes with imputed fixes. An adjusted dataset is constructed of both the successful fixes and the imputed fixes. Inference is then carried out on the adjusted dataset. Weighting relies only on the successful fixes, but weights each fix by the inverse of the probability of the fix being successful. Fixes that have a low probability of being successful are rare and therefore those that are successful have a high weighting, while those with a high probability of being recorded receive a low weighting. Inference is performed on the weighted dataset.

The first method discussed in Chapter 6 is an iterative simulation method based on a model of animal movement. Case studies using synthetic telemetry data and GPS tracking data collected from a single animal are presented in order to demonstrate the method. The second method discussed in Chapter 7 is a weighting method based on a Bayesian Hierarchical Model (BHM). The BHM estimates the detection probability without requiring an explicit solution to the underlying model. A case study using data from a large scale radio tracking study of over 100 animals is used to demonstrate the method and its suitability to large datasets. Both methods

result in a series of nonparametric regression surfaces. Each surface describes the detection probability at a point in time.

5.2 Comparison of regression surfaces

The series of surfaces estimated for detection probability constructed in Chapter 7 show the change in detection probability over time. Chapter 8 and 9 discuss the utility of comparing these surfaces using comparison techniques developed in computational biology. Computational biology is a new discipline which has emerged in the last decade to augment human analysis of medical images using computer programs. Computational biology has developed methods for measuring complex change and detecting anomalies. Two algorithms are presented with adaptations to make them suitable for use on regression surfaces instead of medical imagery.

The first comparison method adapted from computational biology is the Large Deformation Diffeomorphic Metric Mapping (LDDMM) method. The goal of a diffeomorphism is to get a measure of the shape change over time. Diffeomorphic transformation travels along the shortest smooth path between the surfaces. To compare two surfaces, one is designated the source and the other the target. Diffeomorphisms perform a coordinate change and the result is the source surface expressed in the target surface's coordinates. After the coordinate change, the surfaces are then compared. The measures of shape difference used by classical statistics, such as Mean Integrated Square Error (MISE), variance, skew and kurtosis will not be of use in these cases because the space being working on is not Euclidean and the movement is not along straight lines.

There are other application areas where non-parametric surfaces have been proposed to describe variability, but the lack of suitable comparisons has hindered wider adoption. One of these areas

is stock market volatility. Stock market volatility is commonly understood as a measure of investor fear it measures the risk premium built into trading prices in financial market. Cont *et al.* (Cont and Da Fonsecca, 2001; 2002; Cont *et al.*, 2002) proposed volatility surfaces to describe how the volatility of option prices varies for different segments of the market. The segment of the market is the option's location in the market, it is a function of the amount of time before the option matures and the intrinsic value of the option. Volatility surfaces offer more information than the commonly used measures volatility such as the Chicago Board Options Exchange Market Volatility Index (VIX). The VIX and other volatility measures currently traded are weighted averages (Whaley, 2009). These weighted averages collapse the market volatility to an overly simplistic average and mask some of the underlying properties of volatility in different segments of the market. The more informative volatility surfaces have not gained wider adoption because they are not easily compared (Cont and Da Fonsecca, 2001).

Anomalous volatility surfaces are of interest to traders because they signal a change in market behaviour. The Diffeomorphic Demons (DD) comparison method adapted from computational biology is well suited to detect anomalies. DD relies on diffeomorphisms just as LDDMM does. DD is used in this thesis to compare volatility surfaces and look for anomalous conditions of investor fear in stock markets. An index based on the diffeomorphic distance is proposed based on full information from the surface shape with the aim of detecting anomalous trading days where the shape, but not necessarily the level of volatility, is unusual. Both LDDMM and DD have two components, a distance of mapping component and a mismatch component. Estimating the metrics involves minimizing both components. The metrics provide comparisons that echo comparisons made by a subject matter expert. By automating the comparisons a larger number of comparisons can be performed rapidly.

Computational biology has made inroads into the comparison of complex shapes and the inference based of these shape comparisons. The successful reapplication of these techniques is not trivial because the computational method and the new subject matter must be well understood. An assessment of each part of the computational method must be done and if there is not a direct correlation to the new subject matter the computational method either needs adjustment or a different computational method must be selected.

This thesis contains two distinct topics, missing data in radio telemetry studies of habitat use and applying shape measures from computational biology to other subject areas. These two topics are a part of the wider area of spatio-temporal statistical analysis. They are complementary subjects in that the missing data questions give rise to complex surfaces that require a measure of shape change found in computational biology. All of this research arose from investigations into analysis of missing fixes in radio telemetry studies.

6 Estimating automated telemetry system detection probability with animal movement models

6.1 Introduction

This chapter introduces two new methods for correcting bias in Automated Telemetry System (ATS) data. The first method relies on parametric animal movement models which are commonly used in other contexts (Morales *et al.*, 2004) but have not been used for bias correction in any prior work. The second method relies on a non-parametric animal movement model which I developed for this work. Literature for radio telemetry studies and bias correction methods is reviewed, the method is presented and the case study results are reviewed. The case study for the parametric method is a synthetic telemetry dataset which allows checking against known "true" results. The case study for the non-parametric method is GPS tracking data from a study of wild boar conducted in Sweden it is an example of a typical radio tracking study with realistic missed fix issues.

Fieldworkers collecting animal locations can introduce location bias into their observations because animals are more easily detected in some habitats than in others. For example, animals are more easily detected in open terrain than in dense foliage because the foliage can block the fieldworker's line of sight. In addition, some observers are better than others. Radio telemetry studies are one way to reduce the bias introduced by human observers and to study the animals over longer periods of time and greater distances. But just as data from human observers are

biased and should be corrected before analysis can be performed (Buckland, 1993; Sauer *et al.*, 1994), radio devices also introduce bias that should be corrected before analysis.

6.1.1 Review of Literature: Radio Telemetry Studies

Radio telemetry studies collect the location of an animal via a device mounted on the animal and remote radio equipment. The animal's location is determined by a localization scheme such as triangulation, multilateration, trilateration (Newton, 2011). These schemes require the animal mounted device to establish radio communication with three or four remote radio devices. Successful localizations are referred to as fixes. Triangulation, multilateration and trilateration are geometric methods for determining an unknown location (the animal mounted device) relative to other known locations (the remote radio devices). Triangulation uses the geometry of triangles, multilateration uses the timing of radio broadcast signals and trilateration uses the radius of intersecting spheres to fix the unknown location (Newton, 2011). Fixes are collected on a regular time interval. This type of study yields a time series of fixes. Each of the fixes in this time series is then geolocated on a map containing habitat information, to derive the habitat use at the time of the fix. In this way the time series of fixes is converted to a time series of habitat use.

The time series of fixes, and resultant time series of habitat use, can be biased because the variable being studied, habitat, influences the collection of data. The radio transmissions between animal mounted device and remote device maybe compromised in some habitats. Researchers are interested in studying the habitat that the animal uses, but instead of the desired result the collected data represent the habitat that the animal uses which permits radio telemetry fixes. Before the analysis of habitat is carried out the fixes should be corrected to account for this detection bias.

Habitat analysis is often carried out with a Resource Selection Function (RSF). Resource selection functions measure the resources used by an animal as compared to those available to an animal over a fixed period of time. When the use of a resource is disproportionate to its availability use is said to be selective (Manly *et al.*, 2002). The study area is divided into grid squares and each grid square is either used or available. The RSF fit in this chapter is a log-linear model w(x) is of the form:

$$w(x) = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots + \beta_k x_k)$$

Where x_i denote i= 1, …, k independent random variables and $\beta_i$are the selection coefficients. The RSF assumes Poisson counts for the number of used units.

Older radio collar technology introduced location inaccuracies. Locations recorded could be up to 100m from the actual location (White and Garrott, 1990). One source of location inaccuracy was selective availability. Selective availability was the deliberate introduction of error into a percentage of GPS transmissions prior to May of 2000. Correction of location inaccuracy has been well studied (White and Garrott, 1990) and several methods are in common use to remove location inaccuracy (Kufeld *et al.*, 1987; Nams, 1989). Though some intermittent location error is still observed (Villepique *et al.*, 2008), accuracy of GPS has generally improved to a point that it can be ignored because it is within the error of habitat maps (D'Eon *et al.*, 2002).

6.1.2 Review of Literature: Bias correction with iterative simulation

The phenomenon of location bias where the probability of recording a fix is correlated to habitat has been observed in many studies, for examples see Cain *et al. (2005)* and Horne *et al. (2007)*. Simulations studies show that when data with the location bias correlated to the terrain are used

in resource selection models, the resulting models have type II errors for rare habitat use and incorrect conclusions regarding selection versus avoidance (Rettie and McLoughlin, 1999). It is prohibitively costly to collect enough data from stationary collars to quantify the fix rate bias (Johnson *et al.*, 1998). In addition, animal behaviour affects fix rates (Lewis *et al.*, 2007) and therefore bias estimates derived from stationary collars are not directly applicable to moving animals. Other bias correction methods require either observation rates to be introduced *a priori* using additional field data (Horne *et al.*, 2007) or finding time periods in the fix record without animal movement (D'Eon *et al.*, 2002; Frair *et al.*, 2004; Nielson *et al.*, 2009). Additionally, computation with these methods can be time prohibitive for large study areas (Nielson *et al.*, 2009).

The methods for correcting detection bias presented in this chapter are reliant on animal movement models. A range of animal movement models have been developed. Iterative simulation can be used to correct this bias building a model for missing fixes and then repeatedly using this model to impute missing fixes. The imputed fixes are combined with the recorded fixes before analysis (Rubin, 1987; Barnard *et al.*, 2005). The earliest animal movement models were based on random walks where the characteristics of each location or step were independent of the previous steps (Marsh, 1988). A random walk models an animal's movement as a series of random relocations or steps. The length and direction of each step is selected from a suitable distribution and is not dependent on any previous steps making this a memoryless (or Markovian) process. In the next animal movement models –correlated random walks – each step was correlated with the previous steps. This correlation allowed the direction of travel or relocation distance to persist across steps (Kareiva and Shigesada, 1983). More complex models have been developed, such as the state–space movement model (Jonsen *et al.*, 2003; Patterson *et al.*, 2008), behaviour switching through advection–diffusion equations (Grünbaum, 1999) and

mixtures of random walks based on animal behaviour (Morales *et al.*, 2004).

Other factors can introduce detection problems into radio telemetry data. For the remainder of this chapter, the focus is on missing data created by habitat. Static signal boundary and uniform strength are assumed across the study area, and that missing data caused by equipment failure occur completely at random.

6.2 Method

Two methods are proposed in this chapter to adjust for fix rate bias based on an animal movement model: a parametric method and a non-parametric method. The parametric method uses a collection of uncorrelated random walks to simulate animal movement, a method of modelling animal movement proposed by Morales *et al.* (2004). Two distributions are used to express the animal movement: step length and direction. The non-parametric method uses kernel smoothing to create a single non-parametric distribution expressing animal relocations. The non-parametric distribution combines step length and direction into a single distribution. Both methods construct a movement model from the successful fixes, and then use the model to impute the probable locations of missing fixes. In the proposed methods missing fixes are restricted to inside the animal's home range. A model for home range was chosen as the minimum convex polygon that bounds all the successful fixes. The aim of both methods is to model the biological process of animal movement and then use the movement model to iteratively impute the locations for unsuccessful fixes or gaps in the fix record. Incorporating the imputations into the collected data will effectively remove bias from the habitat use data. Note that typically these gaps in the fix record are short, lasting only one to six time steps (D'Eon *et al.*, 2002).

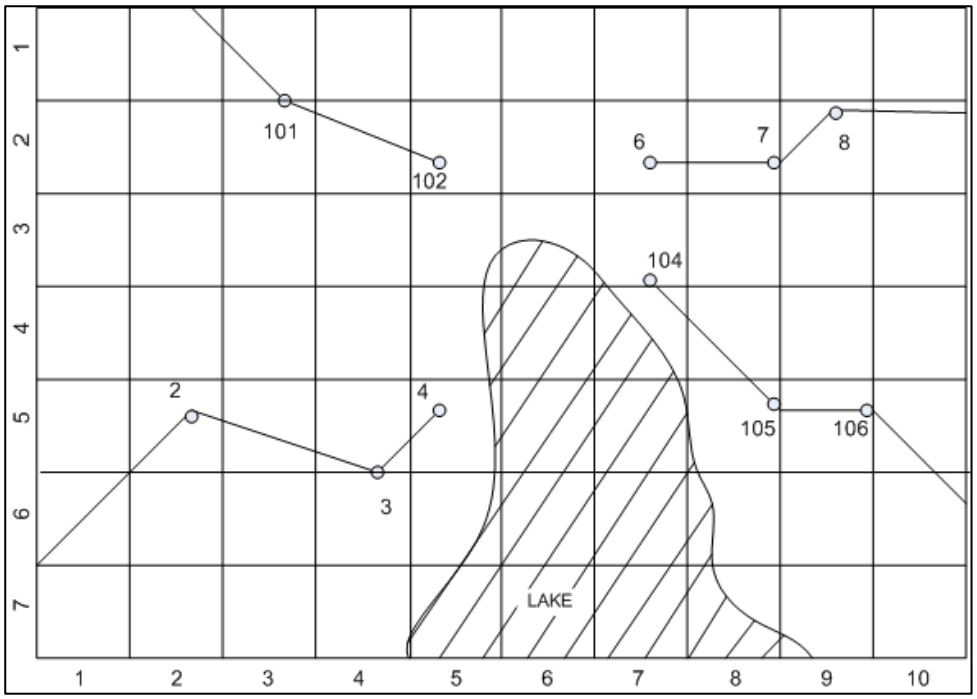

Figure 6.1: Hypothetical set of fixes. Each successful fix is show as an open circle. The area with diagonal shading cannot be accessed by the animal. Fixes 5 and 103 are missing.

Looking at the hypothetical set of fixes in Figure 6.1 where the open circles represent successful fixes, the grid squares with horizontal shading cannot be accessed by the animal, and the fixes at time 5 and 103 are missing. Instead of modelling all the potential paths between time point 4 and 6 and all the potential paths between 102 and 104, these methods will look at each of the cells in the habitat map, (1,1) through (10,7), and use an animal movement model to determine the likelihood of the cell containing a missing location.

The travel distances and directions predicted by the animal movement model and the habitat boundary are used to determine how likely a cell is to be the site of a missed fix. Travel distances and directions are estimated with non-missing data, and then used to make estimates over missing time periods. The more times an animal has missing fixes while in the neighbourhood of the location, the more likely that the location contains missing fixes. Within the model the

habitat boundary ensures that areas that the animal cannot go, e.g. beyond game-proof fences or in lakes, are accounted for.

Instead of focusing on the most likely location of each missing fix, this method reframes the problem and states the expected number of missing fixes in each grid square in the study area. The result of this model is a map of the study area divided into grid squares. For each grid square an estimate is made of the expected number of missing fixes. This result is used to estimate the probability distribution for the location of a randomly observed missing fix. This result is easily mapped and allows straightforward incorporation of habitat covariates such as foliage density and distance to radio receiver.

6.2.1 Parametric Method

The parametric method for correcting the fix rate bias in ATS data models an animal's movement for non-missing periods in the fix record. This animal movement model is then used to iteratively simulate the animal's location during missing periods in the fix record. The animal's movement is modelled as a collection of uncorrelated random walks.

An uncorrelated random walk models is a Markov process comprised of steps or relocations that give the animal's location at regular time intervals. The walk is uncorrelated because during each step the animal is allowed to walk in any direction without regard to the direction of previous steps. The length and direction of each step is selected from a suitable distribution. Because animals display different movement patterns depending on behaviour (Morales *et al.*, 2004) the movement is broken into several simple uncorrelated random walk models. In each of the models, the assumptions of independent time steps and homogeneous animal behaviour hold.

6.2.1.1 Notation

The data consist of a *p*-by-*q* habitat matrix $\boldsymbol{A}$, and a time series *L* with n records, ($L = L_1, L_2 \ldots, L_n$). The matrix $\boldsymbol{A}$ is the habitat available to the animal in the study area. The time series *L* is a record of the animal's location at regular intervals from time $t = 1, \ldots n$. At some time points the animal's location is not recorded successfully this is defined as a missing (or unsuccessful) fix. There are z discrete habitat types in matrix $\boldsymbol{A}$.

The matrix **A** and time series **L** are combined to construct ***U*,** the time series of habitat use. The time series ***U*** gives the habitat use at each time point, $\boldsymbol{U} = (U_1, \ldots, U_n)$, where for each time point U_t is the habitat locations occupied by the animal at that point in time t. When the fix is missing at time *t*, the time series ***U*** is set to $L_t = \emptyset$.

Using the time series $\boldsymbol{L}$, two indicator functions are constructed: the missing function and the endpoint function. The missing function, I_m, indicates for each time *t* if, L_t contains a missing fix. The endpoint function, I_s, indicates for each time *t* if L_t contains a successful fix and one of the adjacent time points, L_{t-1} or L_{t+1}, contain an unsuccessful fix. The missing function, I_m, is 1 when a fix is not successful and 0 otherwise

$$I_m(t) = \begin{cases} 1, & L_t = \emptyset \\ 0, & L_t \neq \emptyset \end{cases}. \qquad [6.1]$$

The endpoint function I_s takes the value 1 for the last successful fix before a missed fix and for the first successful fix after a missed fix, and 0 otherwise

$$I_s(t) = \begin{cases} 1, & L_t \neq \emptyset \; and \; L_{t-1} = \emptyset \\ 1, & L_t \neq \emptyset \; and \; L_{t+1} = \emptyset. \\ 0, & otherwise \end{cases} \quad [6.2]$$

Table 6.1: Hypothetical sequence of locations

Time	Location
2	(2,5)
3	(4,5)
4	(5,5)
6	(7,2)
7	(8,2)
8	(9,2)

Looking at the sequence of locations in Table 6.1, the habitat map in Figure 6.1 and habitat matrix, A

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 3 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 3 & 3 & 3 \\ 1 & 1 & 1 & 1 & 1 & 1 & 3 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 & 2 & 4 & 4 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 & 2 & 4 & 4 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 & 2 & 4 & 4 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 & 4 & 4 & 4 & 4 & 3 & 3 \end{bmatrix}$$

The time series of the animal's locations is

$$\boldsymbol{L} = L_1, L_2 \ldots, L_n$$

$$\boldsymbol{L} = (2,3), (4,3), (5,3), \emptyset, (7,6), (8,6), (9,6)\ . \qquad [6.3]$$

The time series habitat use, $\boldsymbol{U}$, is

$$\boldsymbol{U} = 2,2,2,0,1,3,3\ . \qquad [6.4]$$

The indicator function for missing fixes, $\boldsymbol{I_m}$, is

$$\boldsymbol{I_m} = 0,0,0,1,0,0,0. \qquad [6.5]$$

And the endpoint indicator function, $\boldsymbol{I_s}$, is

$$\boldsymbol{I_s} = 0,0,1,0,1,0,0. \qquad \textbf{[6.6]}$$

6.2.1.2 Determining the correct number of movement models

An uncorrelated random walk movement model was selected because it is computationally simple. In order to simulate complex animal movement behaviours a number of these simple models are combined (Morales *et al.*, 2004). Each time the animal displays different movement patterns, a different uncorrelated random walk movement model is fit. Each movement model is considered a state. The animal can exhibit different states during different behaviours, for example exploring and encampment. Following Morales *et al.* (2004), the animal's movement is subdivided into several random walks, each meeting the appropriate assumptions. Switching between states is not modelled because the current states can be determined either by the successful fixes or by an auxiliary variable such as time of day. For each state a parametric movement model is estimated.

6.2.1.3 The parametric movement model

Once a criterion for subdividing the data has been decided a movement model is fit. Fitting a parametric model requires fitting two independent distributions, one for the step length and a second for bearing. The parametric movement model is based on the uncorrelated random walk movement model. Typical animal movement models of this type have two components, the turning angle and step length. The turning angle is modelled in the typical fashion, but the turning angle is replaced with bearing.

The parametric movement model uses a Weibull distribution for the step lengths. The Weibull distribution has two parameters: a shape parameter, a, and a scale parameter, b. The probability density function is given by the equation

$$P(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} e^{-(x/b)^a} . \qquad [6.7]$$

The Weibull distribution is suitable for modelling the step lengths because it is a right skewed distribution and it is heavy tailed when *a* is small (Zwillinger and Kokoska, 2000). The Weibull distribution is appropriate for animal movement because animal movement is typically modelled as a diffusion process (Blackwell, 1997) and the Weibull distribution is equivalent to simple diffusion in two-dimension when the shape parameter *a* is equal to 2 (Morales *et al.*, 2004). The Weibull distribution is also flexible enough to model more complex situations by letting the shape parameter vary (Morales *et al.*, 2004).

The bearing is modelled either with a wrapped Cauchy distribution or, in simple cases, the circular uniform distribution. If there is no preference for a direction, which would be indicated

by a protrusion in the circular plot, the uniform is appropriate; otherwise the wrapped Cauchy is required. The wrapped Cauchy is the Cauchy distribution wrapped around the unit circle. The probability density function for the wrapped Cauchy distribution is given by

$$P(\theta) = \frac{1-\rho^2}{2\pi(1+\rho^2-2\rho\cos(\theta-\mu))}, \qquad [6.8]$$

where μ is the mean direction and ρ is mean resultant length. The wrapped Cauchy models a preference for movement in a single direction. The circular uniform is the wrapped Cauchy distribution with $\rho = 0$ (Fisher, 1993), thus the probability density function is $P(\theta) = \frac{1}{2\pi}$. The bearing is modelled with the wrapped Cauchy distribution because it has fatter tails than the wrapped Normal distribution. This fat tailed property of the Cauchy distribution is a useful feature for modelling animal turning angles (Wu *et al.*, 2000).

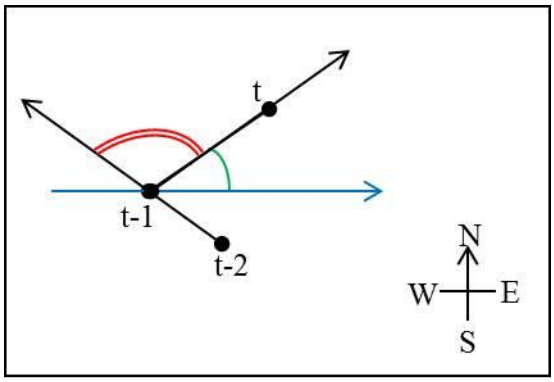

Figure 6.2: Bearing angle vs. turning angle. The bearing angle is shown here in green. It is the angle defined by E (a point due east of t), the location at time t-1, and the location at time t. The turning angle is shown here is red. It is the angle between the trajectory at time t and t-1, and the trajectory at times t-1 and t-2.

The bearing is used in place of the turning angle typically used in random walk animal models. Three points are required to calculate the turning angle, whereas the bearing angle requires only two. Figure 6.2 shows the bearing angle and the turning angle. Both methods measure the

animal's current trajectory by looking at the current position and the previous position. The bearing angle is stated relative to East on a map, while the turning angle is stated relative to the animal's prior trajectory. The prior trajectory is measured using the two previous positions. The bearing angle was chosen to ease computation. The computation is simpler because the turning angle is dependent on the current direction the animal is moving. The current direction of the animal's movement must be calculated at each time point using the prior two fixes, then the turning angle distribution must be rotated to align with the animal's path. The bearing angle is always the same at every location in the study area and the bearing angle has the same orientation as the habitat matrix. Therefore there is no need to calculate the path or rotate the wrapped Cauchy distribution before applying the movement model. This saves two computations each time the movement model is applied.

The proposed parametric movement model is based on a common animal movement model. It is appropriate when the animal shows either no preference in direction or preference to a single direction and the size of the study area and the fix record does not make the computation prohibitively slow.

6.2.2 Non-parametric method

The proposed non-parametric movement model is a novel method for estimating animal movement by recording the location of successful fixes relative to prior successful fixes, then constructing a record of relative animal movements over different time periods. For example, the relative movement of an animal over a one hour period can be modelled or the relative movement of the animal over 5 time steps can be modelled. This record of relocations or steps is then combined into a movement model with kernel smoothing.

6.2.2.1 Movement Model

The non-parametric movement model uses the same notation as the parametric movement model for the habitat matrix $\boldsymbol{A}$, the time series of fixes L, the time series of habitat use $\boldsymbol{U}$, the missing function I_m, and the endpoint function, I_s. See section 6.2.1.1 for full definitions. The non-parametric movement model M is a series of $2n$ simpler models, where n is the largest gap in the fix record L. Each of these models is denoted by a capital M and the subscripts $\pm 1, \pm 2, \ldots, \pm n$, giving the sub-models

$$\mathrm{M} = \begin{cases} \mathrm{M}_{-1}, M_{-2}, \ldots M_{-n} \\ \mathrm{M}_{+1}, M_{+2}, \ldots M_{+n} \end{cases}.$$

The location of each missing fix is estimated with 2 sub models: a forward sub-model with the last endpoint and a backward sub-model with the next endpoint. Given an unsuccessful fix at time t, the prior end point is v, where $\max_v(I_S = 1 \; and \; v < t)$, the forward sub-model is $\mathrm{M}_{+(\mathrm{t}-\mathrm{v})}$, the next end point is w, where $\min_w(I_S = 1 \; and \; w > t)$, the forward sub-model is $\mathrm{M}_{-(\mathrm{w}-\mathrm{t})}$. For example, given a time series of fixes

$\boldsymbol{L} = (12{,}13), (14{,}13), (15{,}13), \emptyset, \emptyset, \emptyset, (17{,}16), (18{,}16), (19{,}16)$ (shown in Table 6.2) the missing fix at time step four is imputed with forward sub-model is M_{+1} using the location of fix 3 and and the backward sub-model is M_{-3} using the location of fix 7.

Table 6.2: Imputations for example L with 3 missing fixes.

Time step	time series of fixes L	Forward imputation	Backward imputation	Endpoint function I_S
1	(12,13)	n.a.	n.a.	0
2	(14,13)	n.a.	n.a.	0
3	(15,13)	n.a.	n.a.	1
4	∅	M_{+1} with step 3	M_{-3} with step 7	0

5	∅	M_{+2} with step 3	$M_{-2}$with step 7	0
6	∅	$M_{+3}$with step 3	$M_{-1}$with step 7	0
7	(17,16)	n.a.	n.a.	1
8	(18,16)	n.a.	n.a.	0
9	(19,16)	n.a.	n.a.	0

Each sub-model $\mathbf{M}_{\pm n}$ needs to be built. A subset of the observations in time series $\boldsymbol{L}$ is used to construct the sub-model. The subset $\boldsymbol{L}_{\boldsymbol{M}_i} \in \boldsymbol{L}$ contains all observations that meet the criteria of model M_i, meaning a successful fix and time t and time $t + i$. A p by-q grid is constructed for each of the $2n$ sub-models, points are placed on the grids using construction functions $C_{\pm n}$ and then kernel smoothing is performed on each grid. The construction function and the locations in time series L are used to place relative fix locations before kernel smoothing. Each pair of successful fixes $\pm n$ timesteps apart is placed on a gird using the function $C_{\pm n}(\theta, d)$ where θ is the angle $\measuredangle E L_t L_{t\pm n}$ where E is a point due east of L_t on the habitat map (or the angle measured on a unit circle placed at L_t) and d is the distance between the points (see Figure 6.3). Using kernel smoothing, at each point L on the model grid the value model is calculated as

$$\mathbf{M}_i(m) = \left(\sum_{\mathrm{l}=1}^{n} k(L_{M_i}, m)\right)\frac{1}{nh^2}, \qquad [6.9]$$

where

$$k(L_{M_i}, m) = \frac{1}{\sqrt{2h^2}} e^{-\frac{d\,(L_{M_i},\mathrm{m})^2}{2}}, \qquad [6.10]$$

where h is the bandwidth set with the normal reference rule, $d\,()$ is Euclidean distance and n is

the number of observations in L_{M_i}. The result is the sub-model expressed as a matrix $\mathbf{M}_i$. The sub-model is the probable location of a fix occurring at time *t*+*i*. The advantage of the non-parametric model is the simplified computation. The non-parametric computation in this application is simply matrix addition.

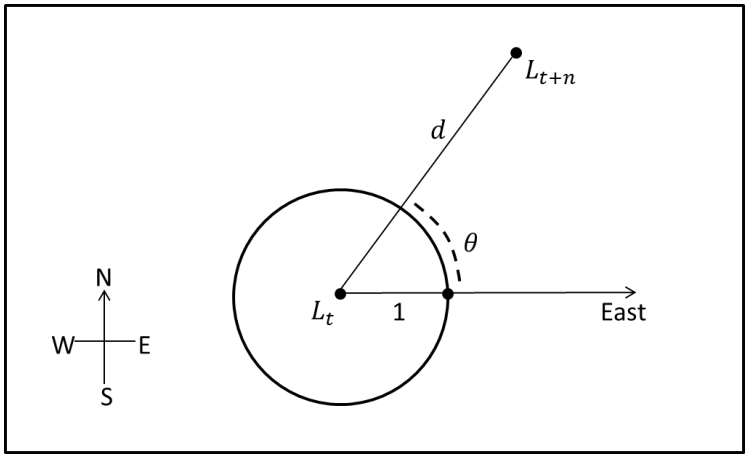

Figure 6.3: Diagram of construction function $C_{\pm n}$, showing the distance d, and angle θ for a pair of successful fixes L_t and L_{t+n}. The unit circle centred L_t is used to measure angle θ.

6.2.2.2 Boundary determination

Confining the travel region allows the inclusion of additional information about the animal's habitat. It prevents placing missing fixes in parts of the study area that either that the animal has never entered or that it cannot enter. The travel region is included via an indicator function that takes the value 1 if a given cell is located inside the allowable range and zero elsewhere. Any home range measure can be included in this way. A common measure of home range, the convex hull (minimum convex polygon), is used here. The convex hull is constructed using all sightings in the study area. Further adjustment is made to remove any regions that the animal cannot use from this convex hull. Other home range measures could be used if they are better suited to the animal's behaviour, e.g. the home range mapping for the animal.

The travel region is stored in a matrix **H** that is the same size as **A**. In the example earlier the travel region is

$$H = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

H is the convex hull containing the animals' sighting contained in Figure 6.1 with the area shaded with diagonal lines removed. This shaded area is a lake and it is removed because the animal cannot swim and is known to be averse to water.

6.2.2.3 Application of nonparametric movement model

The nonparametric movement model is used to determine the likely location of missing fixes. At each endpoint ($I_s = 1$) the sub-models appropriate to the adjacent missing fixes are applied. For example, given the fix record shown in Table 6.2 the 3rd and 7th timesteps are endpoints with three missing fixes between. The sub-models $\boldsymbol{M}_{+1}$, $\boldsymbol{M}_{+2}$ and $\boldsymbol{M}_{+3}$ are centred at the fix location (15,13) and the $\boldsymbol{M}_{-3}$, $\boldsymbol{M}_{-2}$ and $\boldsymbol{M}_{-1}$ (17,16). This has the effect of estimating each missing location twice, once at the prior successful fix location with the relevant forward model and once at the subsequent successful fix location with the relevant backward model.

The landscape and the model are both expressed as matrices, but the matrices are on a different scale. The scale of the model should be a factor of the scale of the landscape. For example, if the landscape is on a 25-metre grid, then a 5-metre grid is a suitable choice for the model because 25 is a multiple of 5. The movement model is on a finer scale than the landscape model so the precision of the GPS fixes is not lost. This can be seen in Figure 6.4 where the finer grid of the model allows the movement model to be centred on the upper right-hand corner of the landscape grid square that contains the last fix. Using a model with the same scale as the habitat map has the effect of relocating each fix to the centre of the habitat grid cell which contains the fix. This is not desirable because the location of the animal is known with more precision than the habitat map.

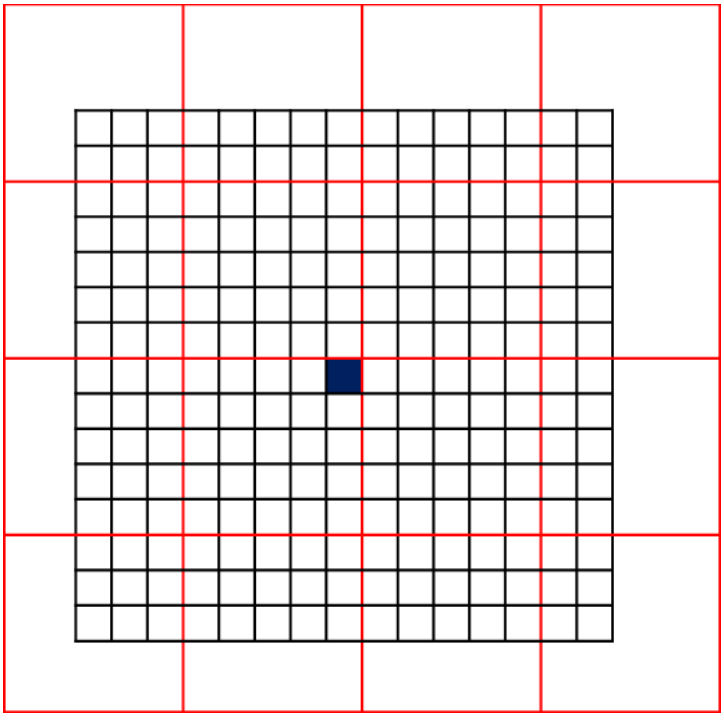

Figure 6.4: The black grid is an illustration of a movement model. The red grid is an illustration of a landscape grid. The blue box contains the last known good fix and the movement model is centred at this point.

To apply a movement model, the method is as follows. Begin with an empty matrix, **B**, which is the same size as the landscape matrix **A**. For every missing value in L, the movement model is centred at the last known and next known fix, and the values of the movement model are added to the appropriate cells of the matrix **B**. After applying the movement model, the entire matrix **B** is adjusted by a normalising constant $2\Sigma_{i=1}^{n} I_m(t)$. Before applying the normalising constant the matrix contains twice the expected number of missing fixes falling in each cell. The final result after applying the normalising constant is the probability distribution for the location of a missing fix, if one were to be chosen at random. This result will be preform simulations and adjust the analysis to account for missing fixes.

6.2.2.4 Adjustment for boundary effect

Confining the travel region creates a boundary effect. An example of this effect is illustrated in Figure 6.6 (b). In this figure, the last known location, L_t, was just inside the boundary. If boundary is disregarded the model will predict locations outside the boundary. If the model is truncated at the boundary, then this missing value will have a smaller contribution to the missing data model, **B**, than missing locations that do not lie close to the boundary.

In order to correct for this boundary effect, the movement template needs to be modified before applying. The model is truncated and the part of the model inside the boundary is renormalized so that it sums to one, just as in the original model. The boundary corrected model is defined as

$$M_i^{adj}(L) = \frac{M_i(L)}{\Sigma_{(L_j)\in M_i} M_i(L)}. \qquad [6.11]$$

This adjusted model can then be applied following the method described in section 6.2.2.3.

6.3 Case Studies

6.3.1 Simulated Data

The first application of the model is to a simulated habitat matrix **A.** The habitat matrix, shown in Figure 6.5 (a), is 100 by 100 and it has two habitat values. The first habitat type (red) increases from 0 to 100 across the rows of the matrix and the second habitat type (blue) increases from 0 to 100 across the columns of the matrix. The animal path was simulated with a preference to both habitat types. The fix record was simulated with the red habitat impeding the signal and the blue habitat not impeding detection. The complete 'true' animal path was simulated with 5,000 fixes and five per cent of fixes were removed as missing.

The animal's path was simulated to prefer both habitats by using an uncorrelated random walk with a Gaussian step length and turning angle weighted to prefer movement towards habitat with higher values of red and blue. The step length used was $N(0,500)$. The turning angle was $\theta = N(0,180)$ measured from the point (2500,0) on the map where both red and blue are maximum. For example, $\theta = 0$ equates to the next step being in the direction of point (2500,0). Simulation of red habitat impeding the fix was achieved by first simulating every potential fix in the animal path, then removing missing fixes. Each fix's success was inversely proportional to the value of the red habitat. The unsuccessful fixes where selected by multiplying a uniform random number between zero and one $rand = U(0,1)$ by the value of the red habitat (0,100) and removing the top 5 % of the results.

The parametric method as discussed in section 6.2.1 was applied to this simulated data. The result is shown in Figure 6.5 b). The dataset was adjusted for missing values by using the results

of the model as a pdf of the missing data. 10,000 Bootstraps estimates were drawn to be used as an estimate of the missing data. The non-missing data was added to each bootstrap estimate to create an adjusted dataset.

a) Habitat Matrix

b) Probable location of missing fixes

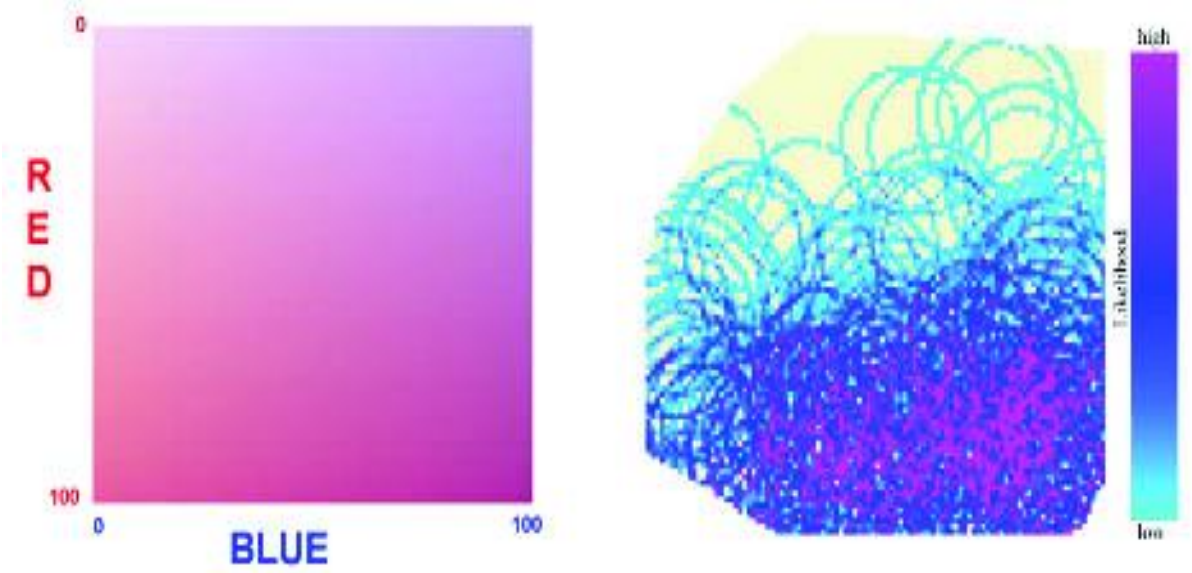

Figure 6.5: a) Simulated habitat matrix habitat matrix 100 by 100 and it has two habitat values and the second habitat type (blue) increases from 0 to 100 across the columns of the matrix. b) The probability of each grid square in the convex hull containing a missing fix.

Using logistic regression, a resource selection function was fit to each adjusted dataset to determine habitat preference for the given bootstrap estimate. The results for the 5th percentile and 95th percentile are the habitat preference for the adjusted data. A resource selection function was also fit to the true data and the data ignoring missing values. A comparison of the three results is shown in Table 6.3. The interpolated values do not change the calculated preference for the blue habitat, which was not simulated to impede detection, but it does change the preference for the red habitat, which was simulated to impede detection.

Table 6.3

Results from incorporating estimates of missing values

Habitat type	True preference	Preference (ignoring missing fixes)	Adjusted preference
Blue	0.62	0.60	0.58–0.63
Red	0.70	0.52	0.61–0.71

6.3.2 Boar Number 1473

The wild boar dataset is comprised of 17 wild boar (*Sus scrofa*, L.) sows tracked via GPS collars in southern Sweden (Thurfjell *et al.*, 2009). A fix was attempted every half-hour. Observations from boar 1473 (named Mia by the researchers) were collected from 14 April 2005 to 26 October 2005. In total there were 9,289 potential fixes, 7,445 (80.1%) of these are successful. The habitat map is the 2006 CORINE Land Cover published by the Commission of the European Communities and covers 36 countries in Europe (Bossard *et al.*, 2000). The CORINE Land Cover Map classifies all habitat larger than 25 ha into one of 44 land use categories. Only six of these land use categories are found in the study area for boar 1473. An extract of the fixes is shown in Table 6.4. The fixes at 02:00 and 03:00 are unsuccessful or missing.

Table 6.4: Extract of fix record for Boar 1473.

time	*Longitude*	*Latitude*	*Locale_N*	*Locale_E*
01:00	13.8539	55.55734	6160376	1376869
01:32	13.85081	55.55666	6160306	1376672
02:00	0	0	0	0
02:30	0	0	0	0
03:00	13.84437	55.55455	6160082	1376259
03:31	13.84447	55.55442	6160068	1376265

The non-parametric method was applied to boar 1473 in the wild boar dataset. The results are

shown in Table 6.5. This boar is active at night and rests in dense foliage during daylight hours (Thurfjell *et al.*, 2009). Two nonparametric movement models were used, one for day and another for night. The data were split into day and night observations based on the sunrise and sunset as defined by civil twilight at the study site (E013 51, N55 33) recorded by the U. S. Naval Observatory. Civil twilight is defined as beginning and ending, "when the centre of the Sun is geometrically 6 degrees below the horizon. This is the limit at which twilight illumination is sufficient, under good weather conditions, for terrestrial objects to be clearly distinguished." (Seidelmann, 1992).

The number of sub-models required for a nonparametric model is determined by the largest gap in the fix record. The largest gap in boar 1473's record was 9 hours (or 18 time steps) during the day and one half hour (or one time step) at night. This means 18 forward sub-models ($\mathbf{M_{+1}}, \boldsymbol{M_{+2}}, \ldots \boldsymbol{M_{+18}}$) and 18 backward sub-models ($\mathbf{M_{-1}}, \boldsymbol{M_{-2}}, \ldots \boldsymbol{M_{-18}}$) are required for the daytime model, while the night-time model requires one forward and one backward sub-model ($\mathbf{M_{+1}}, \boldsymbol{M_{-1}}$). Each of these sub-models is constructed using the record of successful fixes and a function $C_{\pm n}(\theta, d)$ then applying kernel smoothing (as detail in section 6.2.2.1). One of the resulting sub-models is the forward one step night-time sub-model ($\mathbf{M_{+1}}$) shown in Figure 6.6. This sub-model shows the estimated location for a fix at time $t + 1$, using a successful fix at time t. For example, this sub-model ($\mathbf{M_{+1}}$) is used to estimate the location of the missing 02:00 ($t + 1$) fix shown in Table 6.4 because 02:00 is night-time and there is a successful fix at 01:32 (t). The location of the missing 02:00 is also estimated using the backward two step night-time sub-model ($\mathbf{M_{-2}}$) and the location of the 03:00 fix.

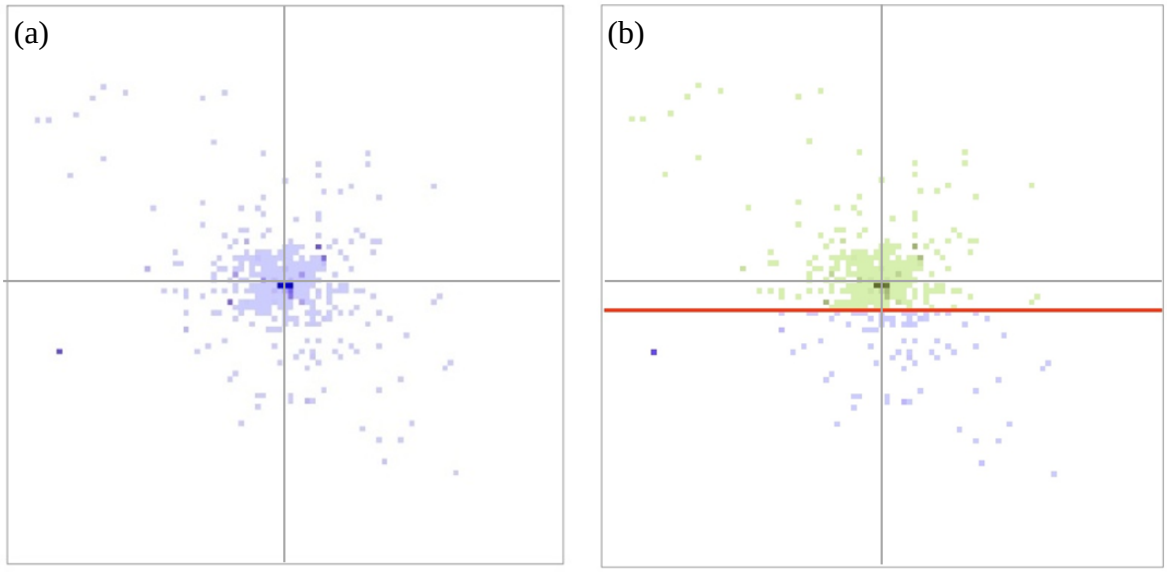

Figure 6.6: (a) The night-time non-parametric 1-step forward sub-model M_{+1} for boar 1473. This model is used to estimate the location at time t + 1, using a successful fix at time t. (b) The same movement sub-model with a boundary. The red line is the boundary, the green part of the model lies within the boundary and the blue part of the model lies outside the boundary.

In the non-parametric 1-step forward night-time sub-model, $\boldsymbol{M}_{+1}$, in Figure 6.6, shows the animal under study moving short distances the majority of the time, but on occasion making large relocations to the NW and the SE. The nonparametric model combines the information in the two parametric distributions (step length or distance travelled and bearing) into a single distribution.

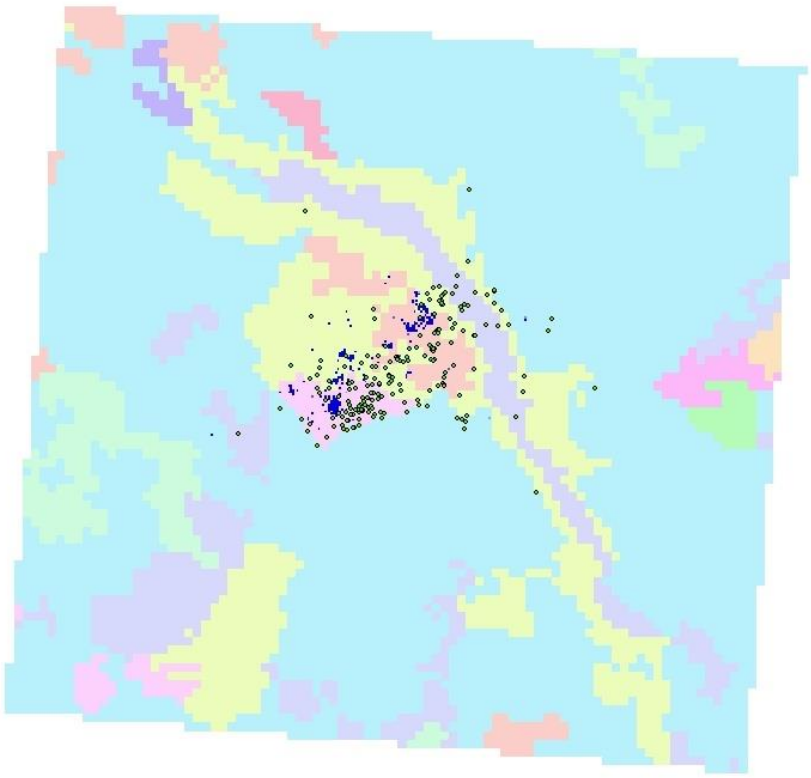

Figure 6.7: Location data for boar 1743. The blue dots represent the non-parametric movement model applied to the landscape. The green dots represent the successful fixes.

Table 6.5: Results bias adjustment with nonparametric model for boar 1473.

Habitat description	Colour in Figure 6.7	Habitat	Recorded	Estimated missing	Adjusted
		———	%	———	
Sport and leisure facilities	Blue	51.18	8.45	14.86	9.66
Pastures	Purple	11.56	2.07	0.29	1.73
Land principally occupied by agriculture, with significant areas of natural vegetation	Green	1.04	0.03	14.63	2.80
Broad-leaved forest	Yellow	27.17	31.77	15.49	28.68
Coniferous forest	Red	5.75	19.14	25.87	20.42
Transitional woodland-shrub	Pink	3.30	38.55	28.86	36.71

A landscape map for boar 1473 with the location of the successful fixes (in green) and the missing fixes as estimated with the non-parametric method (in blue) is shown in Figure 6.7. The adjusted results in Table 6.5 show that pasture land is overrepresented in the collected data before adjustment. This is owing to the fact that pasture land does not impede radio signals with any dense foliage. The majority of the missing fixes are modelled to be in transitional woodland-shrub. This habitat contains dense foliage which impedes radio signals.

6.4 Conclusion

Modelling animal movement allows the interpolation of missing radio telemetry data for a single animal. The missing values are modelled using characteristics of the animal's movement during non-missing periods. The simplicity of the computation allows the adjustment to be carried out on massive datasets.

The parametric version of the model requires fewer computations than other animal movement models to estimate an animal relocation because it uses bearing instead of turning angle. The bearing angle is fixed across the entire study area and the movement direction of the animal does not need to be calculated at each time point. In addition, only a single successful fix is needed to apply the bearing angle calculation, whereas two fixes are needed to calculate a movement direction and apply a model based on turning angle. Because all of the distances and turning angle computations required to apply the parametric movement model are not required by the non-parametric model, the non-parametric model offers computational efficiencies relative to the parametric version of this model. The relative distances between successful fixes are calculated once while estimating the nonparametric movement model. Applying the model to the data is

simply matrix addition. The home range model is also converted to a matrix to speed application. Restraining the model to the home range is done with a computationally simple entrywise product of two matrices. This is important because study areas and the fix record can both be large datasets. For example, the study conducted at the Starkey Experimental Forest which is discussed in Chapter 7 had a study area of over 40 square miles and fixes were collected for over one hundred animals, approximately hourly across more than a decade (Johnson *et al.*, 1998).

Both the parametric and nonparametric models assume that the radio signal strength is constant both across time and space. If GPS collars are used for radio signal then both the transmitter and the receiver are moving; therefore, neither of these assumptions hold true. The model could easily be altered to account for reception in addition to animal movement. Additional future work includes extending this model to include other characteristics of the habitat matrix such as foliage density or distance to the receiver that can affect signal reception.

Modern radio tracking studies offer insight into animal behaviour at small time intervals and large study areas. The bias of the data collected by these studies is recognized, but the methods for adjusting to remove bias are still not widely agreed upon or applied (Frair *et al.*, 2004). The parametric and nonparametric models for estimating detection probability offered here estimate the bias for a single animal directly from the collected data instead of conducting separate studies to determine detection rates of equipment at each habitat type in the study area. These methods provide computationally efficient estimates of missing fixes built on well researched methods of estimating animal movement and animal home range.

6.4.1 Future Work

An area for future work in the area is exploring how to incorporate further animal behaviours into the methods. The difference in boar behaviour during day and night is incorporated by creating two sets of movement models one for day and another for night, but more behaviours could be introduced and state switching based on observations could be incorporated. The model uses each successful fix as an indication of the likely location of the adjacent time points, this is only valid if the time period between fixes is relatively short. In this model, the animal's behaviour is assumed to be the same during the successful fixes and the missing fixes. Some field studies have found that animal behaviour can affect the fix rate (Moen *et al.*, 1996; Edenius, 1997).

7 Constructing regression surfaces for Automated Telemetry System detection probability

7.1 Introduction

This chapter presents a new Bayesian formulation of the Johnson *et al.* (1998) model for large bias correction in Automated Telemetry Systems (ATS). The work builds on the Bayesian Hierarchical Models (BHM) used for bias correction in occupancy estimation and modelling done by MacKenzie *et al.* (MacKenzie *et al.*, 2002; MacKenzie *et al.*, 2005; MacKenzie, 2006; MacKenzie *et al.*, 2006; MacKenzie *et al.*, 2009). There is only one parameter in the BHM process model of MacKenzie *et al.*, detection. When BHM is applied to bias correct ATS there is additional uncertainty requiring correction. I address this additional uncertainty by introducing a parameter, presence, into the process model of the BHM.

In the previous chapter (Chapter 6), a single animal's movement was used to model missing telemetry data. In this chapter, missing telemetry data are modelled for several different animal populations living in the same location. The methods in Chapter 6 could be used to estimate the movement of each individual, but this would be a very tedious process. Instead, in this chapter I present a new method. Looking at the study area as a whole there is information which can be brought to the modelling process from the recurrence of events across animals. For example, multiple animals may have used the same location, or telemetry may have failed for a period on all animals. Using this additional information a large-scale model of the fix-rate bias is built for

a given study area.

Johnson *et al.* (1998) proposed a method for large scale bias correction using the data collected at the Starkey Experimental Forest in Oregon, USA. The Automated Telemetry System (ATS) at the Starkey Forest was used to track multiple species of animals, Elk (*Cervus elaphus*), mule deer (*Odocoileus hemionus*) and cattle, over multiple years. During the course of the study the ATS equipment was improved to the extent that later years of the study offered better fix rates (Johnson *et al.*, 1998). Johnson *et al.* define a stationary time period as six hours where the animal moves less than 200m. These stationary time periods of the fix record are used to estimate the fix rate bias and all other fixes are disregarded for the purpose of bias detection estimation.

MacKenzie *et al.* (MacKenzie *et al.*, 2002; MacKenzie *et al.*, 2005; MacKenzie, 2006; MacKenzie *et al.*, 2006; MacKenzie *et al.*, 2009) have conducted extensive work in the area of incorporating detection probabilities into site occupancy studies. The occupancy estimation models developed by MacKenzie *et al.* leverage multiple site visits to estimate detection probabilities and adjust occupancy estimates for detection rates. This chapter offers a refinement of the method for bias correction presented by Johnson *et al.* (1998) which incorporates non-stationary fixes. The detection bias problem is framed as a Bayesian Hierarchical Model (BHM), drawn on the work of MacKenzie *et al.* where the BHMs developed for site occupancy studies are adjusted for use in bias correction of ATS.

7.1.1 Source of location bias

As discussed in the previous chapter, the use of radio collars to collect information on habitats can be biased when the habitat influences detection (D'Eon *et al.*, 2002; Lewis *et al.*, 2007).

Fixes are made on a regular schedule and sometimes fail. The total number of fixes for an animal at a location is defined to be *f*. These fixes are either successful, f_s, or unsuccessful, U,

$$\boldsymbol{f} = \boldsymbol{f_s} + \boldsymbol{U}. \qquad [7.1]$$

Unsuccessful fixes can occur for many reasons. Links have been made between the fix-rate and many habitat qualities including slope (Lewis, 2007), tree density (Rumble and Lindzey, 1997) and terrain conditions (D'Eon *et al.*, 2002). In addition, many non-habitat reasons for unsuccessful fixes have been established such as animal behaviour (Lewis *et al.*, 2007). The fix-rate, *r*, at a location is the number of successful radio fixes of an animal in that location divided by the total number of attempted radio fixes,

$$r = \frac{f_s}{f_s + U}. \qquad [7.2]$$

There are two prevalent methods for adjusting fix rate bias: iterative resampling and detection weighting (Frair *et al.*, 2010). Iterative resampling replaces missing fixes with estimated fix locations. Detection weighting does not increase the number of fixes, but instead weights each successful fix by the probability of obtaining that fix. The detection weighting, *w*, at a given location is the inverse of the detection rate. Applying the detection weighting corrects for the fix-rate bias,

$$w \cdot f_s = \frac{f_s + U}{f_s} \cdot f_s,$$

$$w \cdot f_s = f. \qquad [7.3]$$

7.1.2 Review of the literature: bias correction with weighting

One technique to estimate fix-rate bias is to place collars in different locations within the study area and then to record the fix-rate. Because each collar is stationary at a location, the fix-rate at that location is the number of successful fixes divided by the number of fix attempts. Field studies that place collars in known locations in different habitat types and directly measure the fix-rate bias have been carried out (Rempel *et al.*, 1995; Moen *et al.*, 1996; D'Eon *et al.*, 2002). Placing collars in each habitat type in a large study area can be very expensive and time consuming, and the results are only applicable to a single brand of collar and the specific study area. Additionally, fix-rates from stationary collar studies do not accurately predict fix-rates in collars worn by animals. This disparity has been linked to animal movement and behaviour (Moen *et al.*, 1996; Edenius, 1997).

A deceptively simple method for removing bias is to use a different temporal scale in the analysis than in the collection. This is done by selecting one successful fix from the analysis period. For example, data collected every half hour can be analysed at the six hour scale by selecting one successful fix in each six-hour period. However, because the missing fixes are biased, the selected fixes may not correlate to the complete record of the animal's habitat use which includes the unsuccessful fix locations. Another problem is the change in temporal scale is a change-of-support issue, and according to Simpson's Paradox (Simpson, 1951), the correlation between variables can change if the support changes. As a result habitat use preferences found by analysing the data at one time scale may not be present at another.

Research into removing bias from telemetry studies continues to grow in conjunction with the growth in GPS studies. Three categories of correction are in use: those based on stationary collars, those based on collars moved by the researcher and those based on collars on wild

animals. Early work in this area was done with fixed and moving collar methods. These methods involve additional field time and money. A large number of factors affecting the fix-rate have been shown in fixed collar studies, including collar height and orientation (Moen *et al.*, 1996), time of day, time of year (Graves and Waller, 2006), tree height (Rempel *et al.*, 1995), density of canopy (D'Eon *et al.*, 2002), battery strength (Gau *et al.*, 2004) and fix schedule (Cain *et al.*, 2005). These findings laid the groundwork for statistical analysis of radio telemetry studies and highlighted areas of further research required to properly analyse animal based collar studies including the need to adjust for location bias and detection bias.

Methods that work directly with collar data from wild animals include Johnson *et al.* (1998), who looked at non-GPS Automated Telemetry Systems (ATS) and developed five linear models for bias correction. Two models use the non-parametric smoothing technique of kriging and three use restricted maximum likelihood (REML). Kriging (or best linear unbiased prediction*)* is a geospatial smoothing technique based on an ordinary least squares estimator (Krige, 1951; Cressie, 1990). Of the bias adjustment models compared by Johnson *et al.*, the kriging methods best fit the example data (Johnson *et al.*, 1998), but all of the models presented by Johnson *et al.* used only a portion of the available data to fit the models. Graves and Waller (2006) proposed sophisticated logistic regression modelling of GPS bias based on habitat properties, satellite placement, and animal characteristics. This approach included comparisons with stationary collars placed in the study area (Graves and Waller, 2006). This method will only work on GPS studies and requires additional data about satellite technology and placement before it is applied to the fix dataset. Frair *et al.* (2004) proposed a logistic regression model. They considered 12 factors: collar brand, time of year, time of day, tree height and breadth, vegetation class, three properties of the canopy, slope, ruggedness and visible sky. The model that best fitted the example data were selected using the corrected Akaike Information Criterion (AIC_c) (Hurvich

and Tsai, 1989). This model utilised factors for collar brand, vegetation class and percentage slope, and an interaction term for vegetation class and percentage slope. After model selection was done the best model was applied to a dataset using both a weighting method of bias correction and an iterative resampling method, and the results were compared and shown to have similar results. Nielson (2009) incorporated this model for detection probability into a method for estimating a Resource Selection Function (RSF) using a detection model. This method relies on the ML estimation methods of occupancy in Mackenzie (2002).

Johnson *et al.* (1998) used a surrogate for stationary collars by identifying periods when the animal was relatively stationary and measuring the fix-rate in these periods. To find stationary periods, they broke the day into four periods each lasting six hours. Because the actual location of the animal is unknown, Johnson *et al.* used the arithmetic mean of the successful fixes during one period as the location. The animal was judged to be stationary if it had at least four fixes in the time period and at most one fix was more than 200 metres from the mean location of the observed animal fixes. They recorded the detection weighting at each stationary site as the total number of fix attempts divided by the number of successful fixes within the period. Not every location had stationary periods, so direct calculation of detection weighting at every grid location in the study area is not possible. The estimation of detection weighting for all grid locations in the study area with stationary time periods is done with five linear models of the form

$$Y = X\beta + \varepsilon, \qquad [7.4]$$

where Y is the vector of observation rates, X is a design matrix of environmental variables, β is the vector or regression coefficients and $\boldsymbol{\varepsilon}$ is the error term.

Johnson *et al.* compare the five models using mean error, model coefficient of determination and

residual plots to select between the models. Johnson *et al.*'s most successful model is the kriging model on a 180 m pixel grid. This model omits all environmental covariates.

7.1.3 Review of the literature: Bayesian hierarchical models

MacKenzie (2002) developed both frequentist and Bayesian techniques for incorporating detection probabilities for animals into large-scale site occupancy surveys. MacKenzie built his models on the assumptions that the area of inference is too large to be surveyed and that detectability is not perfect. MacKenzie's assumptions are confirmed by radio telemetry studies (Moen *et al.*, 1996; Edenius, 1997).

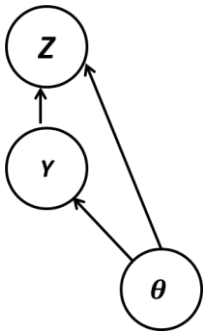

Figure 7.1: Graphical model showing the components of a BHM, where Z is the distribution of the observed data (The Data Model), Y is the scientific process of interest (the Process Model) and θ is the overall Parameter Model.

The Bayesian approach for incorporating detection probabilities in site occupancy is a BMH. Following the terminology from Berliner (1996), there are three levels to a BHM, the Data Model, the Process Model and the Parameter Model. The Data Model describes the distribution of the observed data, Z, given the hidden process $[Z|Y,\theta]$. The Process Model is unobserved and describes the scientific process of interest $[Y|\theta]$. The Parameter Model, $[\theta]$, can be split into data parameters and process parameters $\theta = [\theta_D, \theta_P]$. A graphical model explaining the

interrelationship of the three levels of a BHM is shown in Figure 7.1 Only the data model can be observed, but it is driven by the hidden process model. Both the data and the process models are driven by the hidden parameter model.

The model of interest, the Process Model, is estimated using the posterior distribution as shown

$$[Y|Z,\theta] = \frac{[Z|Y,\theta][Y,\theta]}{\iint [Z|Y,\theta][Y,\theta]dYd\theta}$$

$$= \frac{[Z|Y,\theta][Y,\theta][\theta]}{\iint [Z|Y,\theta][Y,\theta][\theta]dYd\theta}$$

$$= \frac{[Z|Y,\theta][Y,\theta][\theta]}{[Z]} \qquad [7.5]$$

The denominator of the posterior distribution is a proportionality constant that ensures that the total probability is equal to 1. It is dependent on Z and, in large spatial–temporal models, the integral involves thousands of dimensions (Cressie and Wikle, 2011) that make direct computation prohibitive and, in some instances, intractable. One alternative to a direct solution is estimation with the Markov Chain Monte Carlo (MCMC) numerical method. MCMC is a computational method for estimating the posterior distribution; other methods include perfect sampling, rejection sampling, importance sampling and Laplace approximation. For a full discussion of the underpinnings of these computational methods, refer to the seminal works by Wegman (1988), and Efron and Tibshirani (1991). For a resource on implementing computational methods, see Gentle *et al.* (2004) or Martinez and Martinez (2002).

The MCMC method is one of the computational methods made possible by the explosion of computer processing power which has occurred since the 1980s. MCMC derives its power from Monte Carlo integration, which allows calculation of the expected value of the posterior

distribution without knowing the denominator of the expectation function, which is difficult to calculate and sometime intractable, and Markov chains, which generate a sequence from a stationary distribution. MCMC is an iterative method that requires a burn-in period to "forget" the starting values and produce the desired results. After burn-in, MCMC produces samples from the posterior distribution without directly solving it. The posterior distribution can be understood by analysing the samples produced by MCMC.

Following Martinez and Martinez (2002), the posterior expectation function takes the form

$$E[f(\theta)|D] = \frac{\int f(\theta)P(\theta)P(D|\theta)d\theta}{\int P(\theta)P(D|\theta)d\theta}. \quad [7.6]$$

With a change to notation to make it more general, the model parameters are expressed as $\boldsymbol{X} = (x_1, x_2, \ldots, x_d)$ and the posterior distribution as $\pi(x)$. The posterior expectation function can be restated as

$$E[f(\boldsymbol{X})] = \frac{\int f(x)\pi(x)dx}{\int \pi(x)dx} \quad [7.7]$$

Monte Carlo integration estimates samples from the posterior distribution by taking samples from the posterior distribution generated via a Markov chain and calculating

$$\hat{E}[f(\boldsymbol{X})] = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{X}_t) \quad [7.8]$$

where with probability 1, $\frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{X}_t) \rightarrow E[f(\boldsymbol{X})]$ as *n* approaches ∞. A Markov chain generates samples from a stationary distribution, ψ. The success of the MCMC method rests on

this stationary distribution being equivalent to the posterior distribution.

There are several methods for accomplishing this including Gibbs sampling and Metropolis sampling. The Gibbs sampler is computationally less expensive than a Metropolis or a Metropolis–Hasting sampler. Assuming all joint densities are known, the process involves repeatedly estimating $[Y|\theta_P, \theta_D, Z]$, $[\theta_P|Y, \theta_D, Z]$, and then $[\theta_D|Y, \theta_P, Z]$. At each step, the results are updated from the prior step. The resulting stationary distribution is $\psi = [\theta_{D,}Y, \theta_P|Z]$.

In order to use the standard Gibbs sampler the full conditional distributions must be known. If the full conditional distributions cannot be stated explicitly then another sampler is required. This is when it is appropriate to use a subset of Gibbs sampling called the Metropolis sampler, which is computationally less efficient because it discards some candidate samples, and therefore more candidates are required to reach the same sample size. Given a current value of the Markov chain of Y_{cur}, a candidate is drawn for the next value as $Y_{sim}$from a distribution centred at Y_{cur}. With a probability of $min\left(1, \frac{f(Y_{sim})}{f(Y_{cur})}\right) Y_{sim}$ is accepted and Y_{sim} is used as Y_{cur} in the next iteration. With a probability of $1 - min\left(1, \frac{f(Y_{sim})}{f(Y_{cur})}\right)$, Y_{sim} is rejected and another Y_{sim} is drawn. For further discussion on selecting the appropriate sampler and detail on additional samplers, e.g. the Metropolis–Hastings sampler and the independence samplers see Martinez and Martinez (2002). Martinez and Martinez also offer instruction for implementing samplers in MATLAB (MATLAB, 2012).

WinBugs (Lunn *et al.*, 2000) is open source software which performs MCMC sampling. It was written by a team of researchers in the UK and allows users to perform simulation with several different samplers. BHM's can be specified in WinBugs using either graphical models or pseudo-code.

In their site occupancy model Mackenzie *et al.* (2006) use a Data Model from the history of observations at a site, a Process Model of occupancy at each site and a Parameter Model estimated with a vague prior distribution. The Data Model comprises of the number of observations, $Z = (z_1, z_2, \dots, z_x)$ at x sites. The model is expressed as the observed history at each of the sites, conditional on the occupancy of the sites. Given a site i, the latent occupancy variable, z, has two possible states. The states are $y_i = 0$ (the site is unoccupied) and $y_i = 1$ (the site is occupied). For an unoccupied site, discounting false positive detections, the only possible observation is absence. For an occupied site, the probability of observing presence is expressed by a binomial distribution with p being the probability of detection

$$f(z_1, z_2, \dots, z_x | y_1, y_2, \dots, y_x) = \prod_{y_i=1} Bin\,(z_i; p). \qquad [7.9]$$

The Process Model is the occupancy given the latent state variable y_i for the s occupied sites. Each y_i is modelled as an independent Bernoulli random variable with probability ψ giving the equation

$$g(y_1, y_2, \dots, y_x) = \prod_1^s \psi^{y_i}(1-\psi)^{1-y_i}. \qquad [7.10]$$

Thus the posterior distribution is as given as

$$[Z|Y, \theta] = \frac{f(z_1, z_2, \dots, z_x | y_1, y_2, \dots, y_x) g(y_1, y_2, \dots, y_x)}{[Y]}. \qquad [7.11]$$

Vague priors are assumed for p and ψ, a uniform distribution between 0 and 1.

The observations are expressed as a detection history at each site, i. There are two possible

observations at each visit, t, either the animal was observed ($z_{i,t} = 1$) or the animal was not observed ($z_{i,t} = 0$). A short hand notation for a series of site visits is to list the observations in order as a series of zeroes and ones. For example, the detection history $z_i = 1101$ represents four visits to the site. On the first, second and fourth visits, the animal was observed; on the third it was not observed.

The Bayesian techniques for estimating animal occupancy reviewed here offer a framework for incorporating detection. Site occupancy is similar to location bias in that the detection rate is less than 100%, but location bias has the additional feature that the animal is moving. If at time t the animal is detected at location a, then it cannot be detected anywhere else at time t.

7.2 Method

The proposed new method is a refinement to Johnson's model and replaces the linear model with a BHM. The model is used to estimate the weighting function at sites in the study area. It is assumed that at each location repeated samples are made and that over the length of the study, the probability of radio transmission at a given location is constant.

The observations recorded are radio telemetry fixes, with both time and location (Johnson *et al.*, 1998). Following Johnson, the study area is separated into a discrete number of locations by dividing it into grid squares. The spatial neighbourhood of a fix is defined as the grid square containing the fix. Time is separated into time windows of length w. The temporal neighbourhood if a fix at time t is defined as all fix attempts occurring between $t - w/2$ and $t + w/2$.

In the MacKenzie *et al.* (2002) detection rate technique two observations are possible, unoccupied and occupied. In this method the fixes are analysed by temporal neighbourhoods. Given a successful fix, all the fix attempts in the temporal neighbourhood can be found. Each of these fixes is either in the spatial neighbourhood of the fix (f_L), outside of the spatial neighbourhood (f_V), or unsuccessful (U).

The three states can be described with two independent binomial random variables: presence, Y_p, and detection, Y_d. The two states of Y_p are defined to be presence in the spatial neighbourhood ($Y_p = 1$) and absence from the spatial neighbourhood ($Y_p = 0$). The two values of Y_d are defined as a successful fix ($Y_d = 1$) and an unsuccessful fix ($Y_d = 0$). This leads to four possible states: $Y_p = 1 \; and \; Y_d = 1$ the animal is in the spatial neighbourhood and the fix is successful, $Y_p = 1 \; and \; Y_d = 0$ the animal is in the spatial neighbourhood the fix is unsuccessful, $Y_p = 0 \; and \; Y_d = 0$ the animal is not in the spatial neighbourhood and the fix is successful, or $Y_p = 0 \; and \; Y_d = 1$ the animal is not in the spatial neighbourhood and the fix is successful.

Table 7.1: Truth table showing all possible states.

		Observed State		
		In spatial neighbourhood, f_L	Outside spatial neighbourhood, f_V	Unsuccessful fix attempt, U
True State	In spatial neighbourhood, f_L	ρ and ψ	—	ρ and $(1 - \psi)$
	Outside spatial neighbourhood, f_V	—	$(1 - \rho)$ and ψ	$(1 - \rho)$ and $(1 - \psi)$

Table 7.1 assumes that locations are recorded with complete accuracy, and there is no possibility

of accidentally recording a fix outside the spatial neighbourhood when it should be inside the spatial neighbourhood or vice versa. A refinement to the model would be to introduce another source of uncertainty to account for location inaccuracy. Only fixes within a given temporal neighbourhood are considered during this analysis.

The model assumed that at any given time the animal can only be recorded in at most one location. Because the detection weighting method depends on location, the method must decide when an animal is at a given location. The fixed six-hour period used by Johnson *et al.* is replaced by a temporal neighbourhood that acts as a moving time window surrounding the fix. Given a fix time t, all fix attempts occurring between $t - 3$ hours and $t + 3$ hours are defined to be in the temporal neighbourhood of the fix. The parts of the observation record where no fixes are recorded for any animals in the study are discarded. These are considered system wide equipment failure and not detection bias. The truth table in Table 7.1 describes the three possible observed states and the corresponding true states.

The truth table in Table 7.1 is easier to understand in conjunction with an example. Given the hypothetical fixes in Table 7.2 the spatial neighbourhood of the fix recorded 5 is the grid square 1. The temporal neighbourhood of this fix 5 is from 10:00 until 16:00. The highlighted fixes are used to construct the following detection history for the six-hour time window of interest,

$$f_L, f_v, f_v, f_L, U, f_v, U.$$

This detection history states that the animal was detected outside of the spatial neighbourhood for three fixes (the values of f_v in the fix record). The animal was detected in the spatial neighbourhood for two fixes (the values of f_L in the fix record). Finally, the animal failed to be detected two times (the values of U in the fix record). Using the truth table in Table 7.1, a

detection probability is constructed for the temporal neighbourhood of fix 5 as $\Pr(z_5 = f_L, f_v, f_v, f_L, U, f_v, U | Y_p, Y_d)$. The first possibilities are that the animal is in grid square 1 and the fix is successful. The probability of this is $(\rho * \psi_L)$, where ρ is the probability of presence in the grid square and ψ_L is the probability of detection in that grid square. The second possibility is that the animal is present outside grid square 1 and the fix is successful. The probability of this is $(1 - \rho) * \psi_V$, where $(1 - \rho)$ is the probability of presence outside the grid square and ψ_V is the probability of a successful fix outside the grid square. The third possibility is that the animal is present inside grid square 1 and the fix is not successful. The probability of this is $\rho * (1 - \psi_L)$, where ρ is the probability of presence in the grid square and $(1 - \psi_L)$ is the probability that the fix in not successful in the grid square. The final possibility is that the animal is outside the grid square and the fix is not successful. The probability of this is $(1 - \rho) * (1 - \psi_V)$, where $(1 - \rho)$ is the probability of presence outside of grid square and $(1 - \psi_V)$ is the probability of an unsuccessful fix outside the grid square. Looking at the detection history, the last two possibilities cannot be distinguished.

Table 7.2: Hypothetical fix record showing the temporal neighbourhood of fix 5 in grey.

Record	Grid square	Time
1	1	09:00
2	1	10:00
3	2	11:00
4	2	12:00
5	***1***	***13:00***
6	--	14:00
7	3	15:00
8	--	16:00
9	4	17:00

The probability of the observed detection history $(f_L, f_v, f_v, f_L, U, f_v, U)$ in the spatial

neighbourhood of grid square 1 and the temporal neighbourhood of 13:00 is $[\rho\psi_L]^2$ [$(1-\rho)\psi_V]^3[\rho(1-\psi_L) + (1-\rho)(1-\psi_V)]^2$. Assuming that all locations visited by the animal within the temporal neighbourhood have the same detection probability, this simplifies becoming $[\rho\psi]^2$ [$(1-\rho)\psi]^3$[$(1-\psi)$ $]^2$.

In order to estimate the weighting function, MCMC simulation in WinBugs is used. The WinBugs code in Figure 7.2 is used to implement the model described above. The three possible observations in the model f_L, f_V and U are coded as 3, 2 and 1, respectively. Lines 3-6 describe the latent process model for the three possible observations. Line 8 – 17 are the heart of the model and give the conditional probability each the nine possible combinations of observations. Each probability is written in the format pr[ii,<truth>,jj,<observed>] and give the probability of Pr(observed state = <observed> | true state = <truth>). For example, line 16 is the probability of a fix recording a successful fix at site i given that the animal is at site i, $\Pr(z_{ij} = f_L | Y_p = 1)$. Lines 19 – 22 give the prior distributions. Because the number of observations varies at each site the data is setup as two data structures in WinBugs. The first data structure has 73,231 rows (s in the code) one for each site. Each of the rows contains the observations for a site coded as a string of 1's, 2's and 3's. The second data structure (referred to as k in the code) contains the number of observations at each site. The loop started at line 2 iterates through each site (indexed by ii) and the loop started at line 7 iterates through each observation for a site (indexed by jj). The first 10,000 iterations were discarded as burn-in.

```
01: model {
02:  for(ii in 1:s) {
03:    Z[ii,1] <- 1-Yd[ii]
04:    Z[ii,2] <- (1-Yp[ii]) * Yd[ii]
05:    Z[ii,3] <- Yp[ii] * Yd[ii]
06:    Zii[ii] ~ dcat(Z[ii,])
07:    for (jj in 1:k[ii]) {
08:      pr[ii,1,jj,1] <- 1
09:      pr[ii,1,jj,2] <- 0
10:      pr[ii,1,jj,3] <- 0
11:      pr[ii,2,jj,1] <- 1- pr[ii,2,jj,2]
12:      pr[ii,2,jj,2] <- d2[ii]
13:      pr[ii,2,jj,3] <- 0
14:      pr[ii,3,jj,1] <- 1- pr[ii,3,jj,3]
15:      pr[ii,3,jj,2] <- 0
16:      pr[ii,3,jj,3] <- d3[ii]
17:      Observed[ii,jj] ~ dcat(pr[ii,Zii[ii],jj, ])}
19:  Yp[ii] ~ dunif(0,1)
20:  Yd[ii] ~ dunif(0,1)
21:  d2[ii] ~ dunif(0,1)
22:  d3[ii] ~ dunif(0,1)}}
```

Figure 7.2: Model portion of WinBugs code for BHM. The three possible observations in the model f_L, f_V and U are coded as 3, 2 and 1, respectively. The highlighted section of the code is derived from the truth table. The line numbers are for reference only.

This method uses a BHM to estimate a weighting function at every site in the study area. By defining a spatial and temporal neighbourhood for each fix, the probability of a given observation can be calculated. All observations in the telemetry record are used along with the BHM to estimate the detection bias at each location in the study area. The estimate of bias at each site is calculated using MCMC in WinBugs.

7.3 Case Study

This sample dataset is the Starkey Experimental Forest and Range in Oregon, USA. The same dataset was used by Johnson (1998). The 10,102-ha reserve is surrounded by a game-proof fence and includes diverse topography. Elk (*Cervus elaphus*), mule deer (*Odocoileus hemionus*) and cattle were fitted with radio collars and tracked by an ATS. For more detailed information on the ATS, the tracking data and habitat information see the US Forest Service website, http://www.fs.fed.us/pnw/starkey/. The fixes are recorded in the Universal Transverse Mercator (UTM) coordinate system as Northings (UTMN) and Eastings (UTME). The publicly available data for 1993 – 1996 are used in this case study.

The survey area is the portion of Starkey Experimental Forest enclosed by game proof fences. This area is broken into a grid with each pixel being 30 m along each side. A fix location, L, is defined as a single pixel in this grid and the time a fix occurs is denoted t. The BHM model specifies a spatial neighbourhood as grid square containing L and temporal neighbourhood between $t - w/2$ and $t + w/2$. The spatial neighbourhood of a fix at Starkey this is the 30-by-30 metre pixel and the temporal neighbourhood at Starkey is the six hour time window defined by $t - 3$ hours and $t + 3$ hours. Each location, L, has a distinct probability of the animal being at the site ρ_L and a probability of detection ψ_L. The model parameters are estimated with MCMC using Winbugs software. (Lunn 2000); both ρ_L and ψ_L are given vague prior distributions.

Table 7.3: Sample of time points for animal 930318D03. The fixes that are included in the TP for the 21:34:03 fix are highlighted. The locations are recorded in the UTM coordinate system as Northings (UTMN) and Eastings (UTME).

Fix number	Time	UTME	UTMN	Temporal neighbourhood		Spatial neighbourhood (grid number)
				Begin	End	
1	17:29:53	376583	5009976	14:29:53	20:29:53	376455 5010000
2	17:56:59	376590	5009992	14:56:59	20:56:59	376575 5009970
3	18:23:06	376765	5010109	15:23:06	21:23:06	376575 5010000
4	--	--	--	--	--	
5	19:13:50	376917	5010147	16:13:50	22:13:50	376755 5010120
6	19:38:11	377086	5010297	16:38:11	22:38:11	376905 5010150
7	20:04:25	377278	5010275	17:04:25	23:04:25	377085 5010300
8	--	--	--	--	--	
9	--	--	--	--	--	
10	21:34:03	377387	5010343	18:34:03	00:34:03	377265 5010270
11	22:05:20	377215	5010324	19:05:20	01:05:20	377385 5010330
12	--	--	--	--	--	
13	23:08:30	377289	5010368	20:08:30	02:08:30	377205 5010330
14	23:37:18	377345	5010349	20:37:18	02:37:18	377295 5010360
15	00:06:01	377323	5010364	21:06:01	03:06:01	377355 5010360
16	00:36:01	377358	5010322	21:36:01	03:36:01	377325 5010360
17	01:05:37	377480	5010410	22:05:37	04:05:37	377355 5010330

A sample of the data taken from the deer identified as 930318D03 are shown in Table 7.3. The highlighted areas shows the spatial neighbourhood of the fix recorded at 21:34:03 is the grid square 377265 5010270. The temporal neighbourhood of this fix from 18:34:30 until 00:34:03. Following the method described in section 7.2, the highlighted fixes can be used to construct a detection history for the temporal neighbourhood that is the six-hour time window surrounding the fix of interest

$$f_v, U, f_v, f_v, f_v, U, U, f_L, f_v, U, f_v, f_v, f_v.$$

This detection history shows that the animal was outside of the spatial neighbourhood for eight

fixes (f_v), inside spatial neighbourhood for one fix (f_L), and failed to be detected four times (U). The probability of the observed detection history at location 377265 5010270 in the temporal neighbourhood of the 21:34:03 fix is $[\rho\psi_L]^1\ [\ (1-\rho)\psi_V]^8[\rho(1-\psi_L)+\ (1-\rho)\ (1-\psi_V)\ ]^4$. Assuming that all locations visited by the animal within the temporal neighbourhood have the same detection probability, this simplifies to $[\rho\psi]^1\ [\ (1-\rho)\psi]^8[\ (1-\psi)\ ]^4$.

The BHM model has greater variation and higher average weighting than kriging model in Johnson *et al.* (1998). In Figure 7.3 Maps (a), (b) and (c) show that both the kriging model and the BHM model follow the digital elevation model of the study area. Figure 7.3 Map (d) shows the elevation contours in grey and shows the locations where the BHM model gives a higher weighting than the kriging model in black. The BHM model estimates more missing fixes than Johnson *et al.* in areas with low elevation adjacent to steep slopes. This may be explained by the inclusion of non-stationary data in the BHM model.

Table 7.4: Comparison of observation weights

	BHM observation weights	**Kriging observation weights**
Mean	2.33	1.66
Standard error	0.069	0.028
Standard deviation	0.701	0.284
Minimum	1.71	1.33
Maximum	5.30	2.62

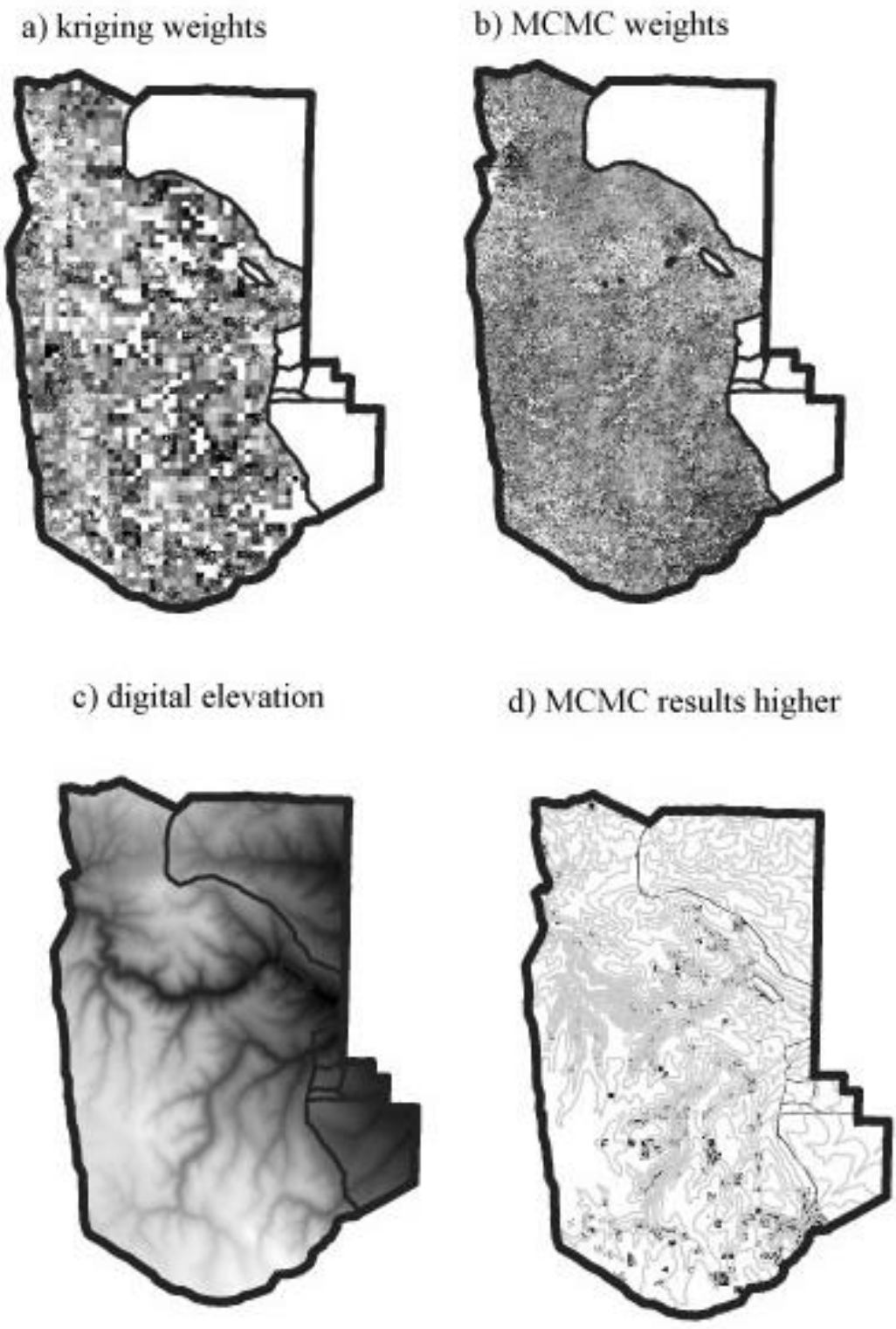

Figure 7.3 : Models applied to the Starkey data. Digital map boundaries, elevation contours, digital elevation model and Johnson's weightings were obtained from the US Forest Service website. In Maps a and b, the shading indicates the fix-rate, where dark grey indicates a low fix-rate and light grey indicates a high fix-rate.

A comparison of the weights from the best performing kriging model (Johnson *et al.*, 1998) and the BHM method are shown in Table 7.4. The vast majority (90.2%) of the kriging weights are in the 95% credibility interval of the BHM. This shows that the only point of difference in the between the kriging model (Johnson *et al.*, 1998) and the MCMC model is in the low elevation adjacent to steep slopes.

7.4 Conclusion

In this chapter a new method for constructing regression surfaces for Automated Telemetry System detection probability was presented. This chapter presented a review of relevant literature in the source of location bias, bias correction with weighting, and Bayesian hierarchical models which were followed by the details of the new method and the results of the method for a case study. The detection probability model presented is an extension of the method for modelling detection probability in Johnson *et al.* (1998). I reframed Johnson *et al.* (1998) using Bayesian methods. This is the first time that a Bayesian formulation of detection probability has been proposed for the correction of ATS bias. The BHM used is based on work done by MacKenzie *et al.* (MacKenzie *et al.*, 2002; MacKenzie *et al.*, 2005; MacKenzie, 2006; MacKenzie *et al.*, 2006; MacKenzie *et al.*, 2009) in the area of occupancy estimation. I expanded MacKenzie *et al.* BHM by adding an additional parameter for presence required to make the model suitable for this context.

Johnson *et al.* (1998) considered one source of variation due to radio transmission and carefully defined stationary periods to eliminate all other sources of variation. A stationary period is defined by Johnson *et al.* (1998) as a six hour window where the animal moves less than 200m.

In this work I introduce a second source of variation, i.e. detectability, and use a BHM. This allows the use of the non-stationary data that Johnson *et al.* (1998) discarded from his model fitting process.

MacKenzie *et al.* (MacKenzie *et al.*, 2002; MacKenzie *et al.*, 2005; MacKenzie, 2006; MacKenzie *et al.*, 2006; MacKenzie *et al.*, 2009) outline a BHM that estimates the occupancy of a site when detection is less than 100%. Because in an ATS study the animal is moving between sites an additional source of uncertainty exists in ATS studies that is not present in site occupancy studies. To account for this uncertainty the additional parameter for presence was introduced into the BHM to make is suitable for estimating the detection probability of ATS.

The method is demonstrated with the same case study also used by Johnson *et al.* (1998), the Starkey Experimental Forest dataset. In this dataset locations for 100 plus animals were collected for multiple years. Datasets of this size are rare and they demand computational efficient methods. The results of the new method are compared to those of Johnson *et al.* (1998). The calculations were done using WinBugs on a personal computer with no special specifications, the estimates the detection probabilities were completed in a reasonable amount of time (less than 4 hours) without expensive computing equipment.

7.4.1 Future work

Future work includes evaluating the need for animal species and habitat covariates in the BMH of detection probability. Further work is needed to understand if different species exhibit distinctly different detection bias in the same location and if explicitly incorporating habitat covariates, e.g. elevation or slope, improve the model. Incorporating additional explanatory covariates into a BHM is a straightforward exercise.

Future work also encompasses incorporating adjacency information into the model of presence using the GeoBugs component of the WinBugs software (Lunn *et al.*, 2000) to incorporate geographical adjacency and allow us to borrow information about detection across sites and make estimations about sites with no recorded fixes. The MCMC probability of an animal being present at a location is currently calculated independently from all of the other location values. Incorporating adjacency information will allow information to be used from the geographic relationship between locations. Johnson (1998) leverages this through kriging. Another area for future work is the improvement of prior distributions used in the MCMC parameter fitting. Instead of vague priors for detection at each site, informative prior distributions based on the signal strength model or the kriging model weights (Johnson *et al.*, 1998) could be used.

8 Comparing regression surfaces with Large Deformation Diffeomorphic Metric Mapping

8.1 Introduction

In Chapter 7 nonparametric regression surfaces were calculated to describe the habitat use and the bias correction weighting of radio telemetry fixes in a study area. This chapter presents a novel use of Large Deformation Diffeomorphic Metric Mapping (LDDMM) (Dupuis *et al.*, 1998; Trouvé, 1998) to compare nonparametric regression surfaces. LDDMM was originally developed to compare medical imagery in computational biology, and this is the first application to environmental science.

Diffeomorphisms are a general class of transformations that operate on manifolds (Trouvé, 1998). A manifold is a smooth surface with either a smooth boundary or no boundary. Diffeomorphisms are used here as measure of the shape difference between the two regression surfaces. The two surfaces are mapped by designating one as the source and the other the target. The diffeomorphic transformation travels along the shortest smooth path between the source and target surfaces. The diffeomorphic mapping can be viewed as a coordinate change where the target surface is expressed in the destination surface's coordinates. Following the coordinate change, the surfaces are compared. There are three restrictions placed on the distance mapping: the mappings must be differentiable, there is a differentiable inverse mapping, and the neighbourhood relationships must be preserved by the mapping.

The measures of shape difference used in classical statistics, such as Mean Integrated Square Error (MISE), variance, skew and kurtosis are not appropriate because the space is not Euclidean and the movement is not along straight lines. In non-Euclidean space, there is no simple formula for distance that is equivalent to the Pythagorean formula giving the shortest distance between two points as a straight line ($d(a,b) = \sqrt{a^2 + b^2}$). The LDDMM metric measures distance along smooth curved paths beginning perpendicular to the source surface and ending perpendicular to the destination surface (Beg *et al.*, 2005). Finding these paths is an iterative process starting with a straight line perpendicular to the source surface. During each iteration, the paths are curved slightly in the direction of the target surface. Each iteration decreases the error between the end of the path and the target surface, the iterations stop when the error stops decreasing with additional iterations (Beg *et al.*, 2005).

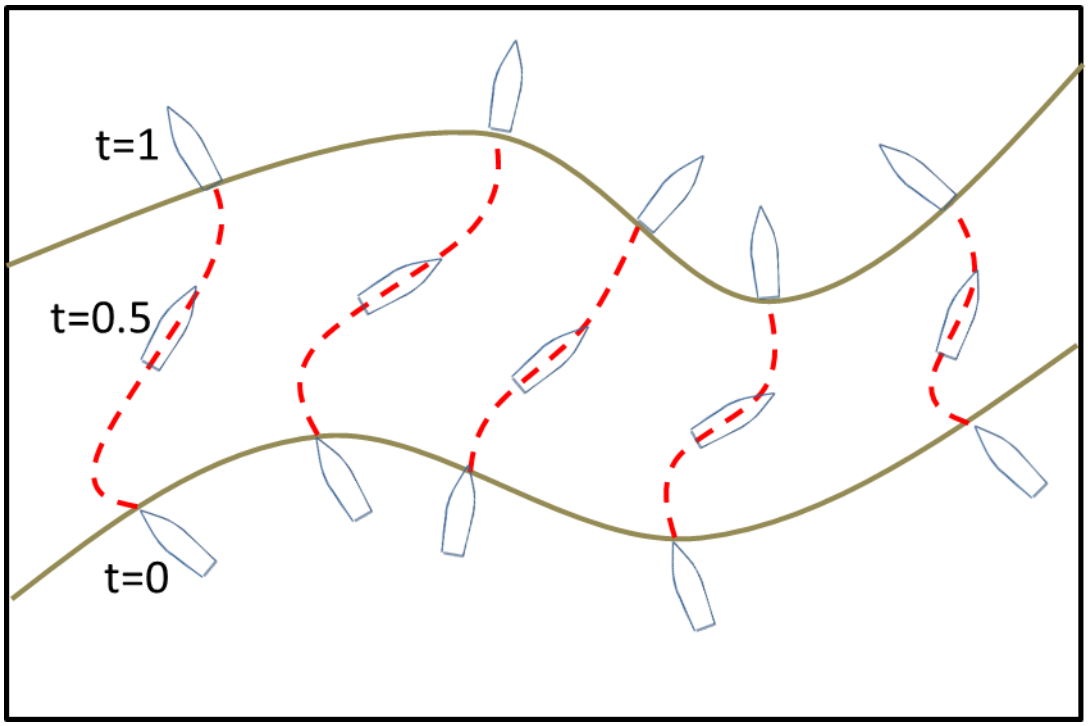

Figure 8.1: Illustration of LDDMM mapping as a fleet of boats sailing from one shore to another. The starting shore is the bottom tan curve and the ending shore is the upper tan curve. All boats start at t=0 and end at t=1. The paths of the boats never cross and each boat starts and ends perpendicular to the shore line.

The LDDMM literature is mathematically dense; I have developed analogies to sailing and driving to aid in understanding. Figure 8.1 illustrates this analogy. Finding the path length of a diffeomorphism is similar to finding a travel time at sea from point a to point b. When setting sail, navigators have to deal with the wind and ocean. In this analogy, the wind is the surface norm and the surface can have varying local shape and orientation, just as the ocean's surface does. The only time the most direct path for sailing from point a to point b is a straight line is if the wind comes from behind the boat and the sea is flat. A diffeomorphism maps a set of points on one surface to their equivalents on another surface. To keep with the sailing analogy, this is analogous to measuring the distance between two shores by sailing a fleet of small boats in a tight formation from one shore to the other. The entire fleet sets sail at the same time and all arrive at the far shore at the same time. There are timing points along the route, for example each boat must cover half of its route by the midway timepoint. Furthermore, no boat crosses another's path and before sailing each boat has its destination marked. The entire fleet sails in a tight formation during the trip while each boat has a different course and speed it must keep its place in the fleet. Each boat must set sail perpendicular to the starting shoreline and land perpendicular to the far shore. Because all boats depart and arrive at the same time, unless the two shores are parallel and there is no wind, some boats will have to accelerate faster than others to cover their route. A boat with a long route will have to accelerate faster than one with a short route to cover the distance in the same time. The average acceleration of the boats is a measure of the distance between the two shores.

Much work has been done performing transformations of this type within the field of medical imagery, giving rise to a new discipline called computational anatomy (Grenander and Miller, 1998). Computational anatomy has proven effective in augmenting human analysis of shape differences in medical images (Beg *et al.*, 2005). LDDMM was first studied by Trouve *et al.*

(1998) and Dupuis (1998) and a computational solution was given by Beg *et al.* (2005). The same techniques are applied here to compare statistical surfaces to augment human analysis of shape difference in statistical surfaces.

Diffeomorphisms are used as a comparison for anatomical structures and they may have use in other places. Medical images collect observations of tissue density at different spatial points. These observations are combined to estimate smooth anatomical surfaces. Similarly, environmental studies estimate the density of items of interest such as animals or pollutants at different spatial locations and use the observations to estimate smooth regression surfaces using nonparametric regression methods. This chapter investigates whether diffeomorphisms can be used to compare abstract structures found in environmental statistics instead of anatomical structures.

Diffeomorphisms are used to compare anatomical surfaces because these surfaces are complex and have so much variation that parametric representations fail to yield useful results (Grenander and Miller, 1998). Despite the difficulty of this task when approached via numerical analysis, even laypeople can observe that all anatomies have a structure and parts of one person's anatomy can be easily mapped to another person's anatomy. A natural mapping can be made of elbow to elbow and wrist to wrist. This mapping allows comparison of two patients. The primary part of computational anatomy focussed on in this research is surface mapping. Surface mapping is used for tasks such as mapping the shrinking within the brain of patients with Alzheimer's disease over time (Beg *et al.*, 2005). These anatomical surfaces are captured with digital imaging technology which produces image files. The image files store information as large matrices in which the values in the matrix represent the density of tissue in the physical structure on a regular grid.

Several similarities exist between medical images and the probability density functions produced by nonparametric and computational statistics methods. Just as in the medical images, these statistical results are expressed with large matrices, where the value of the matrix is a density on a regular grid. In addition, there are identifiable features in the statistical functions. For example, the maxima of two bimodal functions can be matched to each other. If the statistic being studied is spatial (e.g. the population density of an animal species in two study years), the spatial location can be used to help match features and compare the density from one year to the next.

This chapter explores the utility of measuring change in environmental data with diffeomorphisms. Most environmental data are measured at different geographic locations and many nonparametric techniques have been developed to estimate a smooth surface for the variable of interest across the entire region. As in anatomy, non-parametric models have been introduced because parametric estimates have not proven successful, but the comparison of the resulting non-parametric surfaces has proven challenging (Bowman, 2007).

A variety of diffeomorphic algorithms have been developed, but this chapter will focus on Large Deformation Diffeomorphic Metric Mapping (LDDMM). The underlying mathematical framework is discussed in Miller *et al.* (2002), and full details of the computation algorithm are in Beg *et al.* (2005). The LDDMM can be used to measure shape change in volume or in surfaces (Miller *et al.*, 2002; Beg *et al.*, 2005). In order to use LDDMM with non-pararmetric regression surfaces, changes are made to the LDDMM computation to make is suitable for comparing these surfaces. The mechanics of diffeomorphism will be described using an aeroplane dataset for illustration, then, and full details of the computation algorithm are in Beg *et al.* (2005). The LDDMM can be used to measure shape change in volume or in surfaces (Miller *et al.*, 2002; Beg *et al.*, 2005). In order to use LDDMM with non-parametric regression surfaces,

changes are made to the LDDMM computation to make is suitable for comparing these surfaces. The mechanics of diffeomorphisms will be described using an aeroplane dataset for illustration, the method will then be applied to two case studies, the ozone dataset and the Starkey dataset.

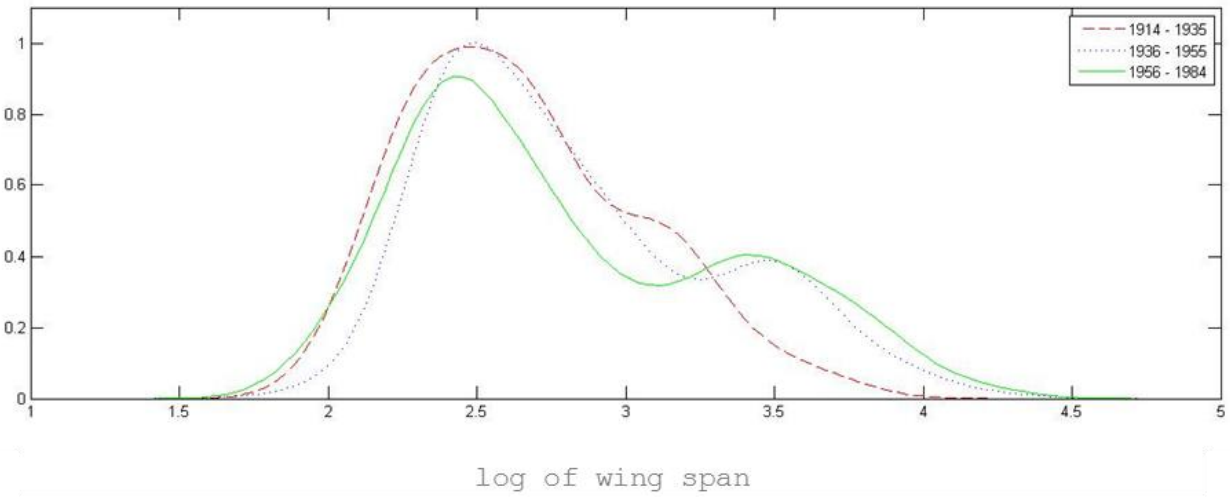

Figure 8.2: A nonparametric density estimation of wing span for production aircraft from 1914 to 1984. Three distinct eras are shown: 1914–1935, 1936–1955 and 1956–1984.

The aeroplane dataset (Wand and Jones, 1995) has been used as a standard dataset for statistical analysis; it describes the evolution of early aircraft design. The dataset consists of the engine power measured in kilowatts, wing span in metres, length in metres, maximum takeoff weight in kilograms, maximum speed in kilometres per hour and the range in kilometres of each production aeroplane from the year 1914 until 1984. The data are divided into three distinct periods, 1914–1935, 1936–1955 and 1956–1984. Regression curves are produce for the wing span in each of the three periods. The curves are shown in Figure 8.2. The resulting bimodal curves are compared. This dataset was used because it is bimodal, it is easily graphed, and it has been used to demonstrate the utility of nonparametric smoothing (Wand and Jones, 1995) It is simpler to graph and explain the mapping and the surfaces in two dimensions, but all the techniques discussed here are intended for higher-order surfaces and mapping. Three-

dimensional datasets, where spatial location is recorded in two dimensions and the third dimension is the environmental measure of interest, have great potential for use.

In environmental studies the two typical comparisons are: comparisons of data from two different time points in the same study area, and comparisons of two different models to describe the same dataset. Comparing two datasets collected at different times in the same study area is analogous to taking a series of medical images to analyse disease progression. The small scale details of the geographic area can change due to changing land use and variation created by the data collection process. The method presented here aligns the two data collections accurately before a comparison is made. Comparing two different models derived from a single dataset is essentially asking the question, how similar are the models to each other? Given the wide variety of methods that can be used to create smooth surfaces, which of the resulting surfaces are the most similar? The method presented here is used to align the two models before a comparison is made. The utility of this method will be shown for both aligning and comparing data collected at different points in time with an ozone dataset and for aligning and comparing two models fitted to the same data with a dataset known as the Starkey dataset.

This chapter will review the current research on comparison of nonparametric regression surfaces, and on surface comparison with computational anatomy. It will detail an original method I developed for comparing nonparametric regression surfaces based on the computational anatomy technique of LDDMM and the application of this method to the two case studies. This will be followed by a review of the results and those of other comparison techniques, and a discussion of further possible research in this area. Because the focus of this research is the utility of the LDDMM algorithm to compare nonparametric surfaces, rather than a study of LDDMM's abstract properties, only methods that can be implemented on a computer

in a reasonable amount of time are presented and in some cases, choices are based on ease of computation.

8.1.1 Review of the literature for comparing nonparametric regression surfaces

The method detailed here relies on Sobolev norms and distances defined by LDDMM to compare nonparametric regression surfaces. This is a valid measure because nonparametric regression surfaces can be expressed in normal decision theory which is defined for Sobolev spaces. Wasserman (2006) shows that nonparametric density estimation can be converted into nonparametric regression. Further, nonparametric density estimation is connected to normal means models following the White Noise Model by Nussbaum (1996). The estimation of nonparametric regression surfaces can be expressed as an estimation of the mean of normal random variables. Because both the nonparametric regression surface and the nonparametric density estimation can be expressed as a normal means problem, they can be analysed with normal decision theory. Normal decision theory is defined for Sobolev spaces and, more generally, Besov Spaces. All of these points support the utility of LDDMM in comparing nonparametric density estimations, as it is possible to interchange regression surfaces and density estimations (Wasserman, 2006).

Wasserman reviews a series of functions which use L_2 distances, which he terms risk and loss functions, to measure the difference between estimated values and measured values. The loss functions are integrated squared error functions of the form

$$L = \int \left(\hat{f}_n(x) - f(x)\right)^2 dx. \qquad [8.1]$$

The risk functions described are integrated mean square error functions of the form

$$R = \mathbb{E}(L). \qquad [8.2]$$

Variations on these functions are derived for different conditions, including Sobolev spaces and, more generally, Besov spaces.

Recently methods for comparing two nonparametric regression curves (or surfaces) have been proposed (Dette and Neumeyer, 2003; Bowman, 2007). These methods focus on comparing regression curves to the underlying observations to determine how many regression curves are needed to model the observations instead of a metric of the shape difference between two regression surfaces. For example, Bowman's method determines if a single curve or a hierarchical set of curves is more appropriate for a dataset (Bowman, 2007). The methods are hypothesis tests with different assumptions and result in a rejection of the null hypothesis or an inconclusive answer.

Prior research on comparing regression surfaces focuses on inference with different hypotheses and different constraints on the datasets. This line of research began with a paper performing inference on the equality of two regression curves (Dette and Munk, 1998). The null hypothesis was that the surfaces are equivalent ($H_0: f_1 = f_2$) and the alternate hypothesis is that the surfaces are not equivalent ($H_a: f_1 \neq f_2$). Subsequent papers use the same hypothesis and extend the research to cover broader assumptions about the data. For example, allowing inference on data where different bandwidths have been used to fit the regression curves or there are heteroscedastic errors in the residuals of the curve fitting (Dette and Neumeyer, 2003). Bowman (2007) introduces an alternate hypothesis of parallel regression surfaces ($H_a: f_1 = f_2 + \alpha,$) for groups of data. The test statistic in all of these papers is the sum of the difference between the

estimate and the observation.

Dette and Neumeyer (2003) proposed two marked empirical processes defined in terms of residuals. The result is a measure of the difference between the observation and the fitted curve. Each of the marked empirical processes is the sum of the absolute error divided by a scaling factor. Bowman (2007) did not use the residuals, but instead used the difference between the regression curves of interest and an additional curve constructed from all available observations. The test used is the sum of the squares of the differences divided by a scaling factor. The aim of both test statistics is to perform inference and determine the number of groups and, consequently, the number of surfaces needed to fit the data.

8.1.2 Review of the literature for diffeomorphisms in computational anatomy

Computational Anatomy is the measurement of shape differences in anatomical structures. For example, if a patient has a yearly brain scan, computational anatomy can measure the change in the shape of the brain over time. If a patient has a heart scan the resulting image can be compared with computational anatomy to the shape of defective hearts and that of a normal heart. The result is two measures, the shape difference between the scanned heart and the healthy heart and the shape difference between the scanned heart and the defective heart. To compare two images, one image is designated as the source image and the other the target, and a mapping is then made from the source image to the target image.

The shape of anatomical objects is difficult to measure because the objects are highly irregular and the instruments used to sample the data introduce systematic bias. Because of this, before measuring shape difference with MSE or L_2 distance, a set of points on the source image is aligned to corresponding points the target image (Grenander and Miller, 1998). A practitioner

identifies geometric features or landmarks in each image. Given these landmarks, all other points will be aligned using an algorithm. The relative amount of "effort" used to align the two brains indicates the distance between them. This alignment process must preserve the relative position of all points.

Following the notation of Grenander and Miller (1998)Following the notation of (Grenander and Miller, 1998), an alignment involves three variables (Ω, H, I), where Ω is the background space of manifolds from zero to three dimensions, $(\Omega \doteq U_\alpha M_\alpha,)$ H is the set of diffeomorphic transformations and I denotes the images. In other words, H maps Ω to Ω, which maps the background space onto itself. An additional parameter, P, is introduced when talking about a population of transformations. P is a probability defined on H, given a population of transformations. P classifies alignments and allows anomalies to be detected.

If the source and destination surfaces meet the smoothness condition of a manifold, a diffeomorphism can be generated to map between them (DuBois *et al.*, 2002). Let us take a simple two-dimensional case as shown in Figure 8.3.

The simplest diffeomorphism maps each point on curve 1 to a point on curve 2 with the same *x*-coordinate. Aligning curves by their *x*-coordinates and then calculating the distance is the method used in other comparisons of nonparametric regression surfaces as discussed earlier.

It is necessary to ensure that the diffeomorphism does not change certain properties of the surface under investigation. For example, the relative position of points on the surface before the mapping should be maintained after the mapping. This is known as preserving the neighbourhood properties (see (b) in Figure 8.4). Points that are not adjacent prior to the transformation have to remain non-adjacent after the transformation. This adjacency rule is

broken if the surface folds back on itself as shown in Figure 8.4 (d). Items that are connected must remain connected, unlike the example curve in Figure 8.4 (c). All of these things would fundamentally change the shape of the surface.

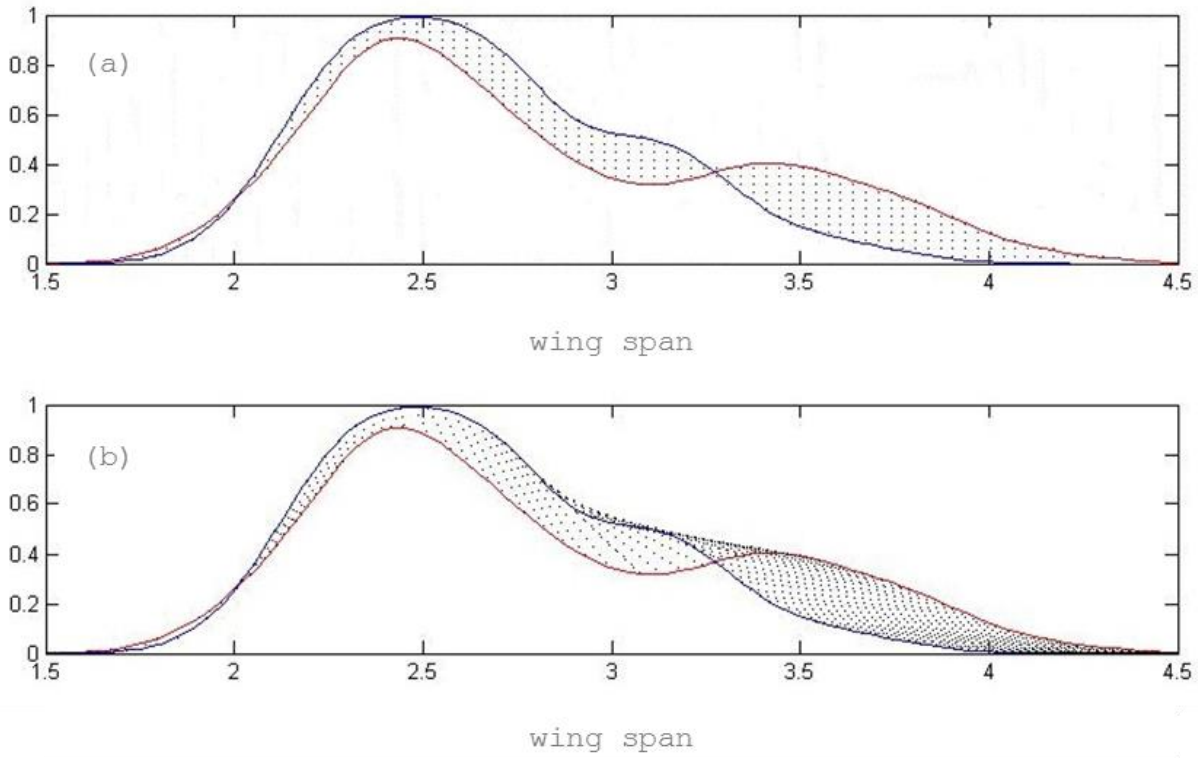

Figure 8.3: Comparison of alignment methods. In (a), each point on the first curve is mapped to the point on the second curve with the same x value. In (b), each point on the first curve is mapped to the point on the second curve with landmarking that aligns the maxima and minima.

It is necessary to ensure that the diffeomorphism does not change certain properties of the surface under investigation. For example, the relative position of points on the surface before the mapping should be maintained after the mapping. This is known as preserving the neighbourhood properties (see (b) in Figure 8.4). Points that are not adjacent prior to the transformation have to remain non-adjacent after the transformation. This adjacency rule is broken if the surface folds back on itself as shown in Figure 8.4 (d). Items that are connected

must remain connected, unlike the example curve in Figure 8.4 (c). All of these things would fundamentally change the shape of the surface.

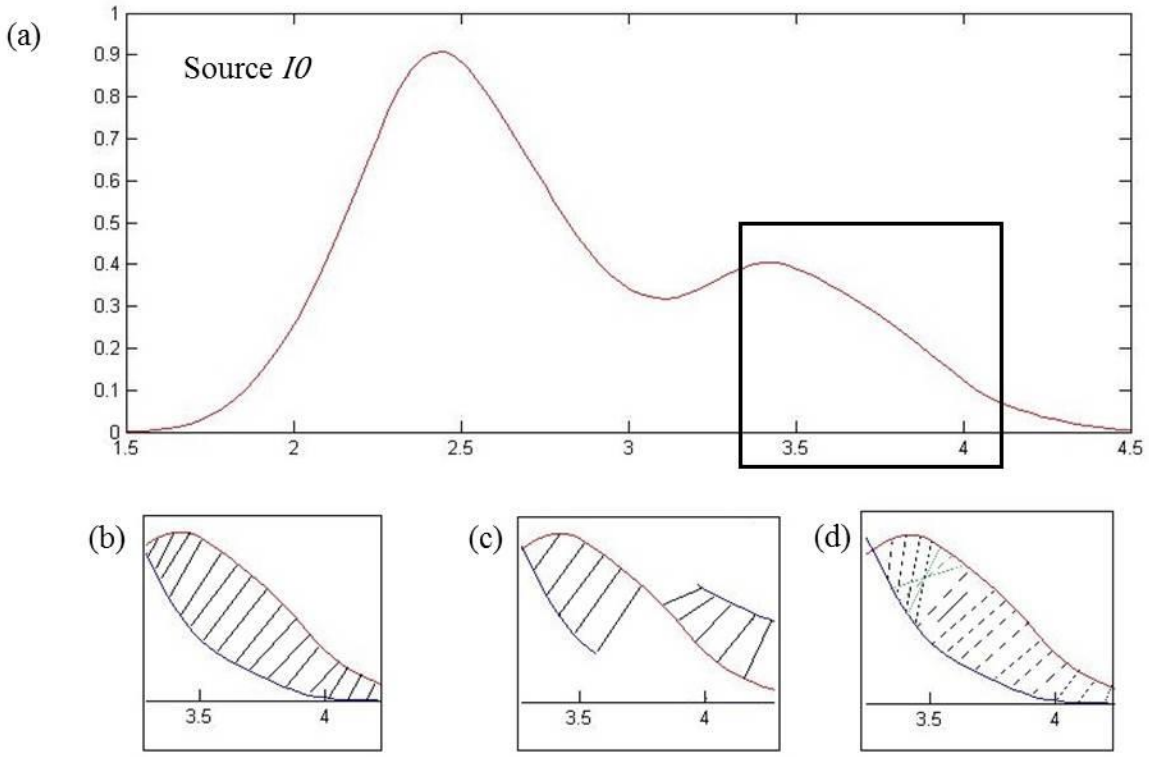

Figure 8.4: (a) Source surface I_0. As shown here, the source surface must be smooth. Close-ups of three different mappings of the inset area showing (b) a valid deformation with neighbourhood preservation, (c) an invalid deformation with points becoming disconnected, and (d) an invalid deformation with the resulting surface folding back on itself.

To ensure that none of these things occur, the diffeomorphic mappings must be one-to-one and onto, and the surfaces must be differentiable everywhere and have differentiable inverses. This means that every point in the source is mapped to a unique point on the target and that every point on the target has a unique preimage on the source. The differentiability of the source and the target surfaces guarantees that the surfaces are smooth.

For any two surfaces, there are many mappings between them that meet the criterion of a smooth neighbourhood preserving mapping, e.g. the two mappings in Figure 8.3. Fortunately, Trouve (1998) shows that there exist a single diffeomorphism that minimises the matching transformation. This minimum transformation is the one of interest, but the only difficulty is finding this transformation.

The computational anatomy metric investigated in this research does not measure the surfaces in relation to the observations or the surfaces in relation to each other, but focuses on the most likely process that could act on one surface and transform it into another. LDDMM is used here as a descriptive statistic and inference is not performed. In order to do inference in computational anatomy, the mapping metric must be defined first. Then the mapping is carried out on a large population and the results are used to form a collection and define the variation of the population so that the empirical probability laws can be defined. Finally, these laws are used to make inferences via Bayesian classification.

Descriptive statistics provide simple descriptors of the shape of parametric curves such as measures of shape (e.g. standard deviation, skew and kurtosis). These measures are used to describe the shape difference between a Gaussian curve fitted to a set of data and the standard normal curve. There is no equivalent description of the shape differences of non-parametric regression surfaces, only the test statistics discussed above for model fitting. Diffeomorphisms provide a measure of shape difference for complex shapes. Even without the benefit of inference, descriptive statistics aid in understanding datasets and the comparison of datasets.

8.2 Method

This section will discuss each in the steps a new method for comparing regression surfaces based on LDDMM and detail the steps developed to arrive at an LDDMM metric for non-anatomical surfaces. These steps are landmarking, discretisation, conversion to image parameter space and finally calculation of the LDDMM mapping. Landmarking aligns the two images so that like features are compared. Discretisation converts the theoretical smooth surfaces to points so that a computer can be used to estimate the results. Conversion to image parameter space expresses the distribution as an image to simplify the amount of coding required. After these pre-processing steps, the LDDMM mapping is estimated.

Figure 8.5 gives an overview of all the components and the notation in use. The diffeomorphism process is estimated as a series of "streamlines" (shown in red in Figure 8.5) that form an estimate of the geodesic curve at a series of points along the surface. Each streamline estimate begins on the source surface and ideally ends on the target surface. Each of the streamlines is comprised of timesteps (shown in dashed blue in Figure 8.5). The constraints on the streamlines are: (1) the transition between timesteps must be smooth, (2) each streamline has the same number of equally spaced timesteps, and (3) each streamline should end at a grid point. It is not always possible for a streamline to end on the target surfaces and be smooth, therefore an error term in introduced (Beg *et al.*, 2005). The error term is the difference between the end of the streamline and the target surface, this error is not shown in Figure 8.5. Furthermore, the first streamline begins on the source surface in the direction of its surface norm and the final streamline ends on an estimate of the target surface in the direction of its surface norm. Many mappings may meet these criteria; the goal is to find the one that minimises both the distance travelled and the error between the target surface and the estimate of the target surface.

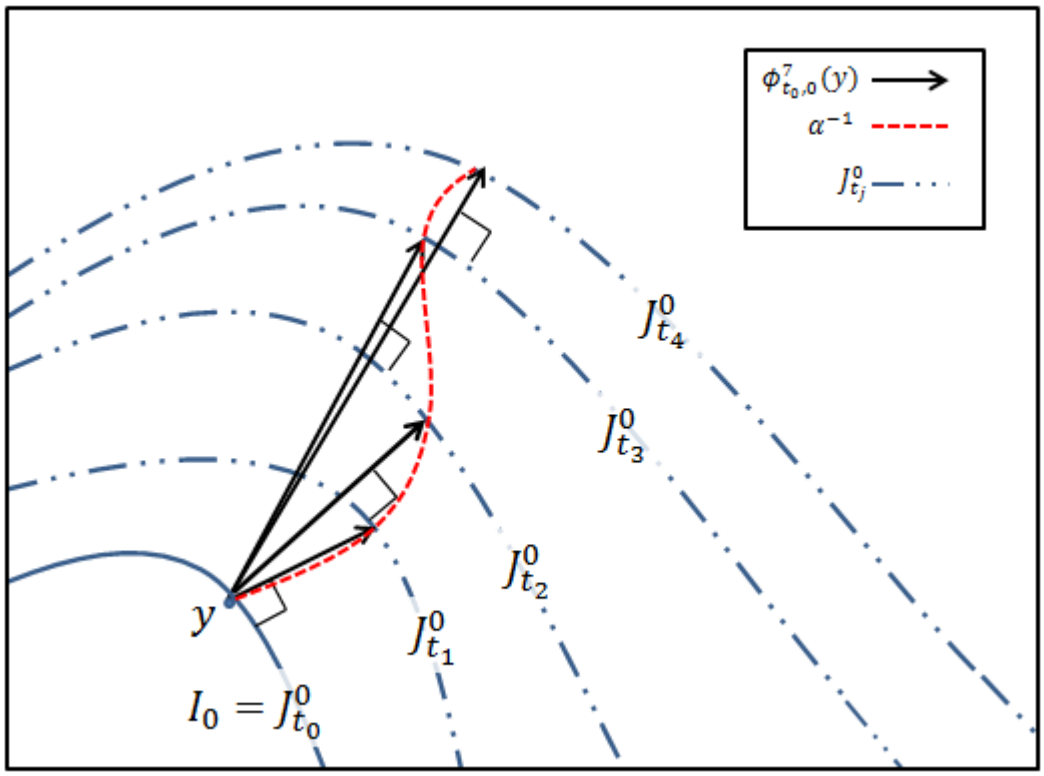

Figure 8.5: Illustration of the mapping of a single point, y. This is the seventh iteration (k = 7) of the algorithm. The interim curve, $J^0_{t_j}$, is shown at each timestep, t_j. The deformation curve is shown in red. The result of the deformation is the curve $J^0_{t_4}$.

It has been shown (Trouvé, 1998) that a single shortest path or "best route" exists. It is not straightforward to find this shortest path because there are no explicit formulae for the surfaces, only a finite collection of points on the surface. In addition, the mapping curve is not defined with an explicit formula. The curve is defined implicitly by estimating an interim surface at each timestep and taking the surface norm. This means that no point mapping can be done in isolation, because the values of the neighbouring points are required to determine the surface norm and therefore the mapping curve. Instead of direct solution, all of the methods presented rely on discrete estimates of the results using computational statistics methods. Beg *et al.* (2005) have shown that the LDDMM algorithm converges to the best path if run for sufficient time. In practice there is no way to know when a sufficient time has passed and convergence has

occurred (Beg *et al.*, 2005). The computational estimate is made by starting with an initial estimate of the path, iteratively improving the estimate, and stopping when there is no improvement over prior estimates. Given a long enough run, this is the best estimate given the starting criteria.

Following the notation of Beg *et al.* (2005), the number of timesteps is T, the number of points on the surface is N, and the size of the timestep is dt, giving $T = N \times dt$. Each diffeomorphic curve is denoted with three indices: j denotes the origin point on the initial surface ($j \in [0, N]$), t_j denotes the timestep ($t_j \in [0, T]$), and k denotes the estimate number ($k = 1,2, \ldots K$). Using these indices, the forward mapping is denoted $\alpha_{t_j}^k(y)$ and the corresponding velocity is $v_{t_j}^k(y)$. The starting point of the mapping is J_0^0, which is defined as the starting image (I_0) and J_T^0 is the end point. The flow is also estimated at T timesteps, also beginning at the target image, I_1, and J_0^1 is the end point. This mapping is denoted as $(\alpha_t^v)^{-1}$. The curve defined by the mappings, the starting surface, the ending surface and the interim surfaces are all smooth. The metric mapping is based on the mapping curve.

8.2.1 Estimating a smooth surface

Anatomical transformations begin by creating smooth surfaces from observations in the form of medical images. Similarly, statistical transformations begin by creating a smooth surface with kernel density estimation for a set of observations. Another description of nonparametric regression is infinite parameter regression (Wasserman, 2006). Kernel density estimation is a nonparametric estimate of the probability density function (PDF) for a set of observations. Given a random d-dimensional sample $X_1, X_2, \ldots, X_n$, the continuous PDF, f, is defined as

$$f(x) \geq 0, \int_{\mathbb{R}^d} f(x)\, dx = 1, \qquad [8.3]$$

where

$$Pr(a \leq x \leq b) = \int_a^b f(x)dx. \qquad [8.4]$$

The function is required to be smooth and it cannot be represented with a finite number of parameters.

Kernel density estimation was first described by Rosenblatt (1956). The theory has been extended by many authors since. An extensive summary of the evolution of this field can be found in Izenman (1991). Izenman notes that "nonparametric density estimation has become a prominent statistical research topic," and he points out that nonparametric density estimation can describe datasets that contain multiple modes more accurately. Several methods are used for nonparametric density estimation, including Maximum Penalised Likelihood, Average Shifted Histogram and kernel methods. This research will focus exclusively on the kernel methods. Both Wand and Jones (1995) and Hastie and Tibshirani (1990) provided comprehensive treatment of the use of kernel methods. Izeman's notation and treatment of the topic has been utilized throughout this chapter.

With a few exceptions (Fryer, 1976; Deheuvel, 1977), the research in this area has focused on large-sample properties. The goal of all nonparametric density estimates is to provide unbiased, consistent estimates, i.e. bona fide density estimates. The definition of an unbiased estimator, $\hat{f}$, for a function f, is one that for all $x \in \mathbb{R}^d, E_f[\hat{f}] = f(x)$. Unlike the case of parametric densities, it has been shown that no single nonparametric density estimate can be unbiased for all

continuous densities (Rosenblatt, 1956). To overcome this, sequences of estimators $\hat{f}_n$ that are asymptotically unbiased for f are used. The consistency of a distribution ensures that the estimate $\hat{f}$, converges to the actual distribution, f. There are several types of consistencies: the (weakly) pointwise consistency, the strongly pointwise consistency and consistency defined through error criteria such as the L_1 and L_2 approaches discussed in Hall (1989). Finally, any estimate that satisfies $f\ (x)\ >=\ 0$, $\int_{\mathbb{R}^d} f\ (x)\,\mathrm{dx}\ =\ 1$ is considered bona fide.

The multivariate kernel density estimator is defined as

$$\hat{f}_{h(x)} = \left(\mathrm{nh^d}\right)^{-1}\Sigma_{\mathrm{j}=1}^{\mathrm{n}}\,\mathrm{k}\left(\frac{X-\mathrm{X_j}}{h}\right),\ x\in\ \mathbb{R}^d, \qquad [8.5]$$

where k is the kernel function and the window width is defined as $h\ =\ h_n\ >\ 0$. Scott (1992) showed that this estimate, $\hat{f}$, is equivalent to a histogram with an infinite number of bins. Prior to Cacoullos (1966), k was refered to as the weight function. The estimate $\hat{f}$ is constructed by centring a kernel at each observation. The simplest kernel is a function that satisfies the property $K(x)\geq\ 0,\ \int K\ (t)dt\ =\ 1$.

Using a kernel that meets this criterion ensures that the estimate will be bona fide. The PDF is estimated from the kernels by averaging all the kernels at each point in the domain. Because the area under each kernel is 1, the average of the areas under all the kernels, $\hat{f}$, will also be 1. The kernel function can be chosen to be any function meeting this criterion. The Gaussian kernel with unbounded support is used as follows

$$K\,(t)\;=\;1/\sqrt{\,(2\pi)}\;exp\;(-\,{t^2}/{2}).\qquad[8.6]$$

The optimal kernel has been shown to be $K(t) = \frac{3}{4}(1 - t^2);\ -1 \le t \le 1;$ and 0 otherwise (Epanechnikov, 1969). The optimal kernel is not used here because that the normal kernel simplifies the calculations and has similar efficiencies to the optimal kernel (Silverman, 1986; Scott, 1992).

The second parameter that must be chosen is the bandwidth. It is generally understood that the choice of bandwidth is much more important that the choice of kernel (Marron and Nolan, 1987). The optimal bandwidth is the bandwidth that minimises the Asymptotic Mean Integrated Square Error (AMISE)

$$AMISE_{ker}(h) = \frac{R\,(K)}{nh}\;+\;¼\,\sigma_k^4\,R\,(f''),\qquad[8.7]$$

where the kernel K is a continuous PDF with $\mu_k\;=\;0$, σ_k^2 is the variance K, $0 \le \sigma_k^2\;<\;\infty$ and $R(g)$ is a measure of the roughness of the function given by $R(g) = \int g^2(x)dx$. The corresponding optimal bandwidth is given by

$$h^*_{ker}\;=\;\left(\frac{R(K)}{n\,\sigma_k^4\,F(f'')}\right)^{1/5}.\qquad[8.8]$$

This holds under the conditions described in Parzen (1962) and Scott (1992). The optimal bandwidth cannot be used because σ_k and f'' are unknown.

Because a Gaussian kernel is used here, the normal-reference rule can be used to estimate the

optimal bandwidth. The normal reference rule is $h^*_{ker} = \left(\frac{4}{3}\right)^{1/5} \sigma$, where σ is estimated by $\hat{\sigma}$ as either the interquartile range divided by 1.348 ($IQR \,/\, 1.348$) or the standard deviation, which always produces the smallest window (Silverman, 1986).

Because several surfaces are being estimated and compared, it is important to ensure that the comparison measures differences in the surfaces and not in the bandwidth. For this reason, the same bandwidth will be used for all the surfaces. The bandwidth is calculated for each surface using the normal reference rule, and then taking the average of all these bandwidths with the formula

$$\hat{h}^*_{ker} = \frac{\sum h^*_{ker}}{s}, \qquad [8.9]$$

where s is the number of surfaces being estimated. This average of the bandwidths is used for all surfaces.

Figure 8.6 shows an example of the resulting nonparametric estimate of the probability density function on the three different time periods. This estimate has some elegant statistical properties. Using the L_1 approach to consistency, Devroye (1983) showed that for any K where $K(x) \geq 0$, $\int K\,(t)dt = 1$, $\hat{f}$ will be a strongly consistent estimator of f as the bandwidth approaches zero ($h_n \to 0$) and n times the kernel approaches infinity ($nh_n^d \to \infty$) as n approaches infinity ($n \to \infty$). No conditions are required for f.

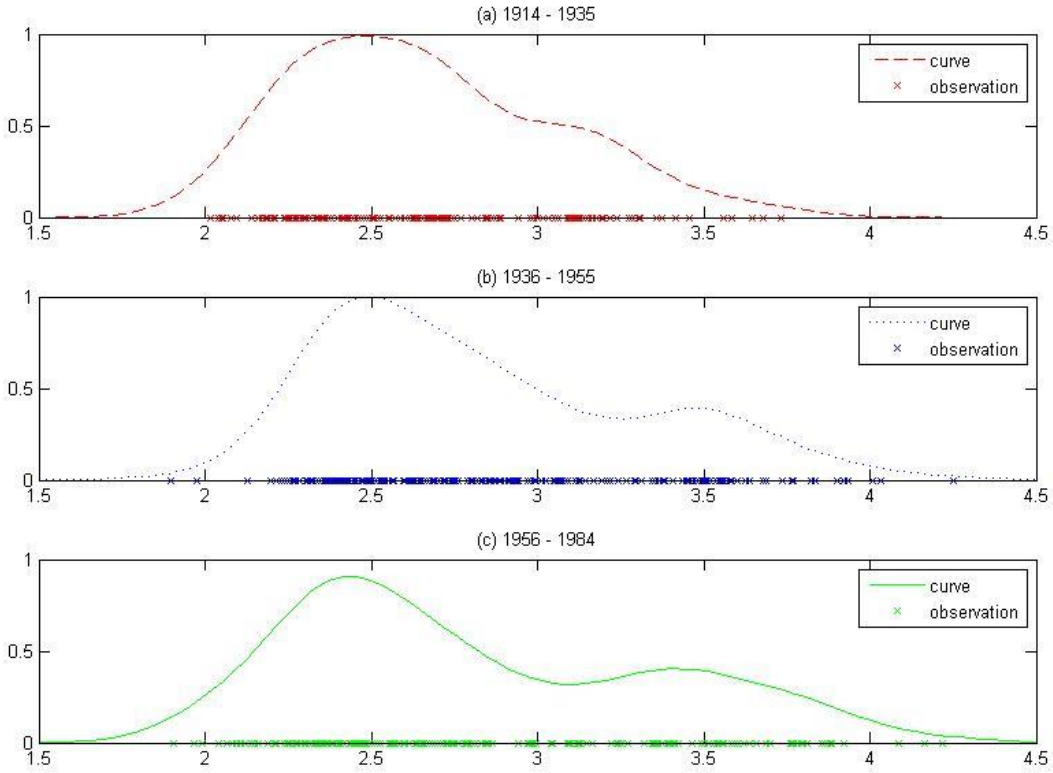

Figure 8.6: Nonparametric density estimates for wingspan over three periods in aircraft design. The observations are shown as crosses on the axis.

In the univariate case, Mean Integrated Square Error (MISE) has a very low computational cost, lower than that of a histogram (Devroye and Penrod, 1984). Using the L_2 approach, several consistency results have been shown (Hall and Hannan, 1988). One result shown for the univariate case is that given the condition $h_n \rightarrow 0$ as n approaches infinity, the estimate is asymptotically unbiased and asymptotically normal (Parzen, 1962).

In conclusion, the smooth surfaces estimated by kernel estimation methods provide consistent, asymptotically non-biased, bona fide estimates of the PDFs; therefore comparisons of these surfaces represent comparisons of the PDFs.

8.2.2 Identifying landmarks

In computational anatomy, landmark transformation is a coarse mapping of the template image to the target image. The theoretical concept underlying landmarking can be traced to work by Kendall (1984). The idea of using a coarse landmarking of an object was applied to identifying functionally distinct areas brains, for. For example, the Allen brain atlas that identifies 1000 anatomical landmarks on the adult human brain (Holtz, 2011). These landmarks are identifiable by subject matter experts. The diffeomorphism uses the landmarks to perform and guide the mapping.

This thesis is the first application of LDDMM on non-anatomical objects. In anatomical settings, manual landmarking is not always performed before LDDMM is performed, but some other type of simpler alignment method is always applied prior to LDDMM. LDDMM is classified as a high-dimensional large-deformation image-matching transformation. Lower order transformations include the affine transformations and rigid transformations. Affine transformations preserve colinearity and orientation. They are defined as mappings from $\mathbb{R}^n \rightarrow \mathbb{R}^n$ that can be written in the form $F(\boldsymbol{p}) = A\,\boldsymbol{p} + \boldsymbol{q}$, where A is a linear transformation for all $\boldsymbol{p} \in \mathbb{R}^n$ (Gallian, 2010). Rigid transformations are a subset of affine transformations comprised of only rotations and translations. Anatomical mapping is carried out through a series of mappings of progressively higher dimensions, beginning with rigid affine transformation and working through progressively more complex transformations, the last of which would be in the class of LDDMM (Grenander and Miller, 1998). For example, Beg *et al.* (2005) remove rigid rotation and translation prior to LDDMM using "Analyze" software (Robb and Mayers, 1996).

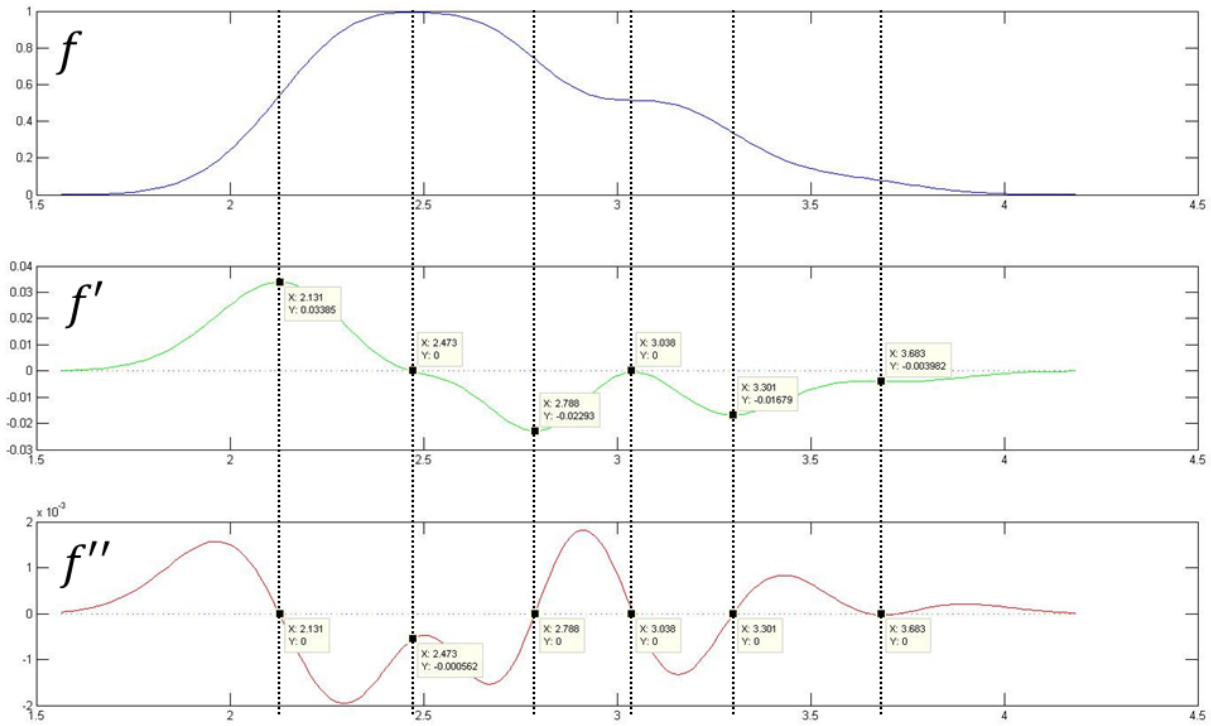

Figure 8.7: The 1914–1935 curve of the aeroplane dataset which shows how landmarking was performed. The first and second derivative with maxima and minima are used as landmarks to align this curves with others.

Each step in the anatomical application of LDDMM was paralleled in this application. A landmarking process was selected for use on statistical surfaces because the features of one distribution are mapped to corresponding features in the target distribution. For this dataset, functionally distinct areas of the nonparametric probability functions to be mapped must be identified. Figure 8.7 shows landmark values for the surface *f*. There are six landmarks on the curve ($L = 6$), specifically, these landmarks are $l_1 = 2.131,\ l_2 = 2.473,\ l_3 = 2.788,\ l_4 = 3.038,\ l_5 = 3.301$ and $l_6 = 3.683$. In this case, the points of interest were selected as landmarks to align the curves: the local maximum, the local minimum and the points of inflection. This was natural choice because these values are easily calculated. Assuming that the function is twice differentiable, the local maximum and minimum, and the points of inflection

can be identified as follows,

If $f'(x) = 0 \; and \; f''(x) < 0$, x is a local maximum;

If $f'(x) = 0 \; and \; f''(x) > 0$, x is a local minimum;

If $f'(x) = 0 \; and \; f''(x) = 0$, x is a possible inflection point.

This method was then used to perform a coarse mapping of I onto I'. Note that I and I' must have the same number and type of extrema. Each image is expressed as a series of landmark points, $I_0 = \{x_l \in \Omega, l = 1,2,\ldots,L\}$ and $I_1 = \{y_l \in \Omega, l = 1,2,\ldots,L\}$. Then the corresponding landmarks are mapped $x_l \rightarrow y_l$, for $l = 1,2,\ldots,L$.

In this way, a rough mapping is made from one regression surface to the other, thus ensuring that the resulting mapping will match similar areas on each surface. If there is an underlying landscape, as in the case of environmental study areas, the landmarking can be done on physical landmarks such as a moving tree line during periods of deforestation.

8.2.3 Surface discretisation

So far, this chapter has explored the theory of calculating a smooth surface and the method for a coarse mapping of point correspondence between these surfaces. The computations are performed with a computer and calculations are done using a computational grid. This computational grid is used to discretise continuous functions. Before discussing the computation, discretisation and other computational considerations required to perform these calculations will be discussed.

Often, the theory for both nonparametric statistics and computational anatomy is discussed using integrals, limits and large sample properties. Computers are used to do a large number of discrete calculations to estimate these continuous structures. It should be noted that in some literature, a discrete representation for a continuous surface is referred to as a "surface parameterisation". Instead of this term, "surface discretisation" is used to avoid any confusion with parametric statistics.

Table 8.1

Step	Grid or background space	Computation	Distance within grid	Distance between grids
1	Fixed regular Cartesian mesh	Kernel smoothing	Euclidean	Area between curves or displacement
2	Unstructured surface grid	Surface discretisation	Arc length	–
3	Unstructured surface grid relative to reference surface	Landmarking	–	–
4	Image grid	Image conversion	–	–
5	Another image's grid	Diffeomorphic change of coordinates	–	LDDMM and mismatch

Table 8.1: Summary of computational grids and their distance measures

Table 8.1 shows the five steps required to compare two surfaces and the associated computational grids. The first discretisation estimates the smooth surface. This method follows the standard practice of estimating the value using a fixed regular Cartesian mesh (Wand and Jones, 1995). This mesh has a consistent distance between each point as measured along the axis; therefore Euclidean distance can be measured on this mesh. Using this mesh, the surface is estimated at unit intervals along the axes. The unit interval is fixed across all the surfaces being

estimated.

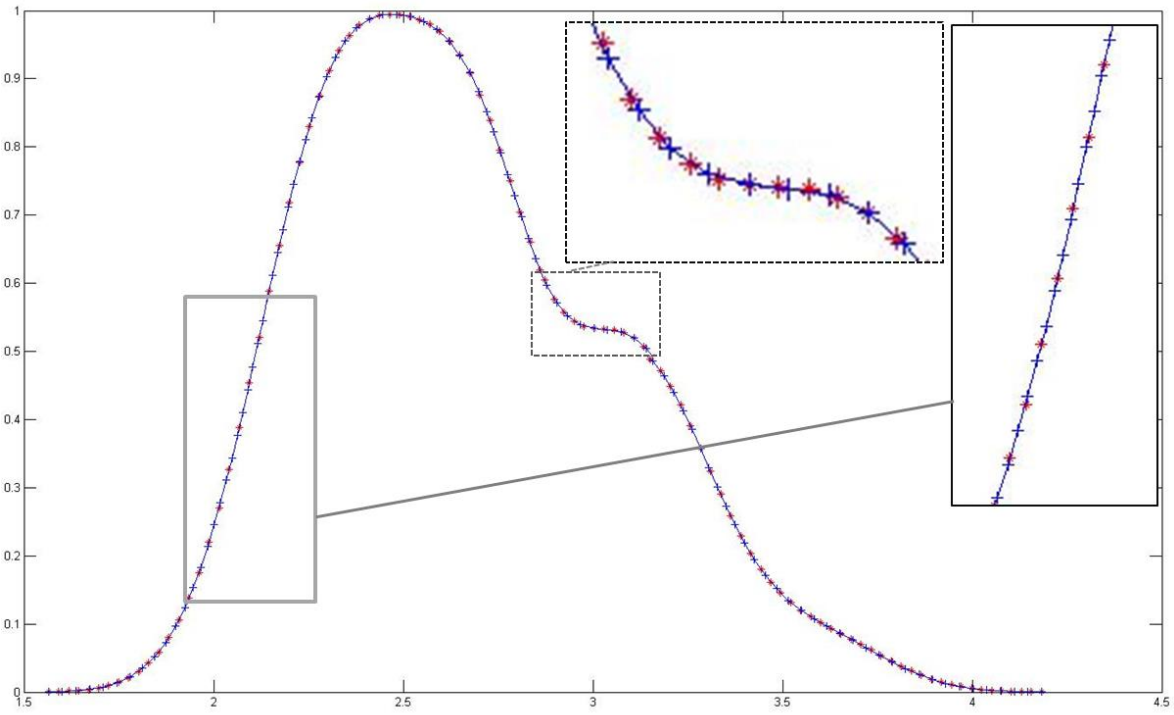

Figure 8.8: A comparison of surface parameterization with x-parameterization. The red circles are equally spaced along the x-axis and the blue crosses are equally spaced along the surface of the curve.

8The diffeomorphism measures changes in the surface, and therefore a surface-based computational grid is required. This is one of the class of grids known as unstructured grids (Staniforth and Cote, 1991). One common surface based grid used for anatomical surfaces is the triangular grid. For example, the cortical surface of the brain is represented as a smooth two-dimensional surface of class two, C^2 (Joshi *et al.*, 1995). The surface is produced by mapping each point in the three-dimensional space ($\mathbb{R}^3$) to a point in two-dimensional space ($\mathbb{R}^2$). Given the axes of $\mathbb{R}^3$ (x, y, z) and the axes in $\mathbb{R}^2$ (u, v). The mapping of a point p on the surface D is $x(p) = \{x(u,v), y(u,v), z(u,v)\}^t\ (p) = \{x\ (u,v), y\ (u,v), z\ (u,v)\}^t \in \mathbb{R}^3, p \in D \subset \mathbb{R}^2$. In this computational grid, distance is measured along the surface.

A similar unstructured grid was created based on the surface coordinates for the smooth surfaces created by density estimation. This is done to obtain a grid with fixed distance between points as measured along the curve instead of along the axes. The benefit is that the rapidly changing area of the surface can be accurately represented. Because the spacing of the Euclidean mesh relies only on the value of the x coordinate, this mesh under-represents places in the curve where the y-values change more quickly than the x-values. Another computational advantage of the surface mesh is that fewer dimensions are needed to store points. For example, a three-dimensional surface with three axes, x_1, x_2, y, can be flattened into a two-dimensional surface and thus represented as a two-dimensional matrix. Similarly, a curve in two dimensions, $x\ and\ y$, can be expressed as a one-dimensional line where each point has a single coordinate representing the distance from the origin. In order to do this surface discretisation, the origin must be defined as a point on the surface to be flattened.

Changing from the Cartesian coordinate system to a surface coordinate system forces the measurement of distances, angles and relative position in terms of the surface. There are several approaches to this change in coordinates, with semi-automated discretisation via landmarking being selected here. Line discretisation involves defining an origin and a single parameter, u_1, which gives the distance along the surface of the line to any other point on the line. More formally, any point on the line $x \in D$ can be uniquely represented as $x = X(u_1)$ for some parameter space $\boldsymbol{u} = (u_1) \in D \subset \mathbb{R}^1$. Initially points equidistant along the x-axis are assigned on each surface. In order to change coordinates, the required distance measured along the curve must be computed. This is done by first fitting a parametric spline approximation to the set of points, then integrating to get the arc length. In order to place points at equal intervals measured on the surface, an iterative process is carried out on the points that are equally spaced by x-distance. During each iteration the points are adjusted and the arc length between points is again

calculated. When the arc lengths between all points are equal, the process is stopped (see Figure 8.8). The steps as are as follows:

(1) Using the mesh based on equally spaced x-coordinates, fit a parametric spline.
(2) Integrate along the parametric spline and get the approximate arc length between each point on the mesh.
(3) Using the entire arc, calculate the desired unit length (the length of entire arc divided by the number of points in the mesh).
(4) Find new x-coordinates so that intervals greater than the desired length are decreased and intervals less than the desired length are increased.
(5) Using the new x-coordinates, fit a parametric spline.
(6) Repeat steps 2–5 until the arc length between each set of points is equal to the unit arc length.

The result is a mesh equally spaced by arc length.

The LDDMM method relies on the mechanics of fluid flow. The theory of LDDMM is established with mathematical proofs using continuous variables, then the solutions are estimated using computer-intensive methods (Beg *et al.*, 2005). Two essential components for arriving at an accurate estimation are choosing a grid which accurately reflects the background space of the problem and a robust method of estimating change over time at the points on this grid. The computational grid and the discretised time points are the main sources of calculation error. Calculation errors can swamp the values being estimated. LDDMM involves searching for minimum values; if the calculation errors are too large, then the estimate will find the minimum calculation error instead of the desired result. A review of various change estimation considerations is found in Staniforth and Cote (1991).

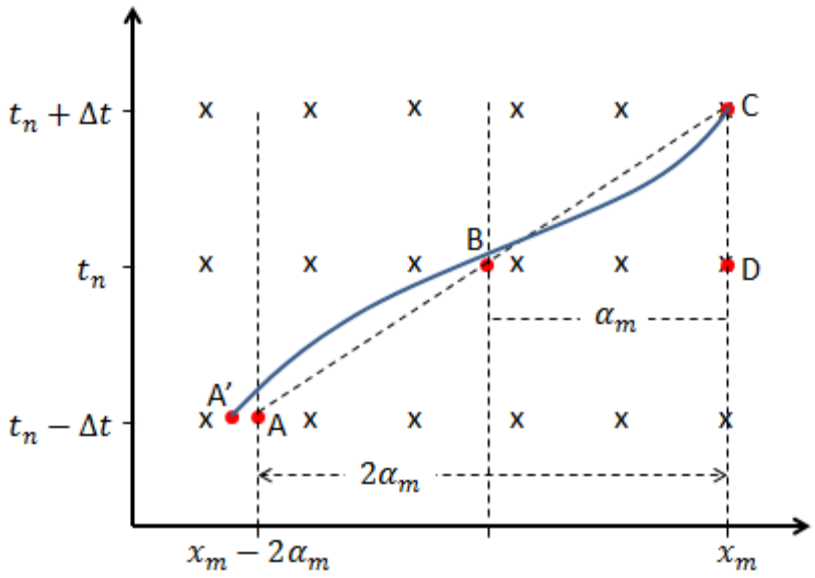

Figure 8.9: Iterative backward two-step trajectory. The crosses are the computational grid. The dotted line between A and C is the estimate of the actual trajectory shown by the solid blue line between A' and C. Adapted from Staniforth and Cote (1991).

This chapter follows Beg *et al.*'s (2005) implementation of LDDMM and uses a two-step semi-Lagrangian scheme with an iterative backwards trajectory to minimize computational error. The error introduced by calculating particle movement with a grid is illustrated in Figure 8.9 which is an adaptation of an illustration in Staniforth and Cote (1991). The figure shows one-dimensional movement over time, the grid points are shown as crosses, the horizontal axis gives position and the vertical axis shows time. The solid line is the actual trajectory of the point that arrives at position x_m at time $t_n + \Delta t$ and the dashed line is the estimate of this trajectory. The particle has moved a distance of α_m in the x direction at time Δt. While the actual curve begins on grid point A', the estimate begins slightly away from this grid point at A. This means that between time points the same particles are not seen. At time $t_n + \Delta t$, the particle at point C is the particle that arrives at that point at that time. This same particle may not have been on a grid point at time t_n. The semi-Lagrangian method of propagating fluid flow over time is a compromise between the Eulerian method, which operates on a fixed grid (Robert, 1981), and Lagrangian methods, which

track the location of regularly spaced particles over time (Staniforth and Cote, 1991). The drawback of the Eulerian method is that is forces very small timesteps in order to stop errors from building up (Robert, 1981). The drawback of the Lagrangian methods are the results are in an unstructured grid with poor estimation at some areas in the surface, stemming from the fact that the initially regularly spaced particles tend to clump up over time (Welander, 1955). The semi-Lagrangian method selects different particles at each time point. The particles are selected to arrive exactly at Cartesian coordinates, thus minimising the problems of both prior methods.

8.2.4 Conversion to image parameter space

The next grid used for calculation is the image grid (step 4 in Table 8.1). The grids in step one to three used real numbers, either $\mathbb{R}^2$or $\mathbb{R}^1$. This grid is discrete. Each surface will be converted to a TIFF image (Figure 8.10), because most image manipulation software can accept TIFF images. Using already written image manipulation software simplified writing the code.

TIFF images are stored in three-dimensional matrices. Each mesh point on the grid is converted to a pixel. TIFF images are comprised of pixels (the pixel is the smallest unit of an image, typically visualised as a small coloured square). The storage of a pixel requires three integers, with each integer being in the range of 0–255, $p \in 0,1,2,\ldots,255$. The three cells taken together represent the amount of red, blue and green respectively, (p_r, p_b, p_g). For example, the pixel $(p_r, p_b, p_g) = (255,0,0)$ has 100% red, 0% blue and 0% green. Surfaces are converted into greyscale images. Greyscale images are TIFF images with all three colours (red, blue, green) set to the same value for each pixel, $\forall p, p_r = p_b = p_g$. In conversion of surfaces to TIFF images, the value of 0 and 255 on the TIFF scale are mapped to the minimum and maximum values of the surfaces, respectively. The minimum value in the tiff matrix is $\min_{TIFF} = \min_{\mathrm{p}\in\mathrm{D}}(\mathrm{p})$ and

the maximum value is $\max_{TIFF} = \max_{\mathrm{p}\in\mathrm{D}}(p+\delta)$. The mapping is restricted to increments of $1/256$, $p_{TIFF} = \left\lfloor \frac{p-\min_{\mathrm{TIFF}}}{\max_{TIFF}} 256 \right\rfloor$. The pixels are combined into a matrix, with one dimension being the number of pixels in the height and the other being the number of pixels in the length of the TIFF, the third is the three grey scale values of each pixel, the result is a three dimensional matrix that holds the values of the image.

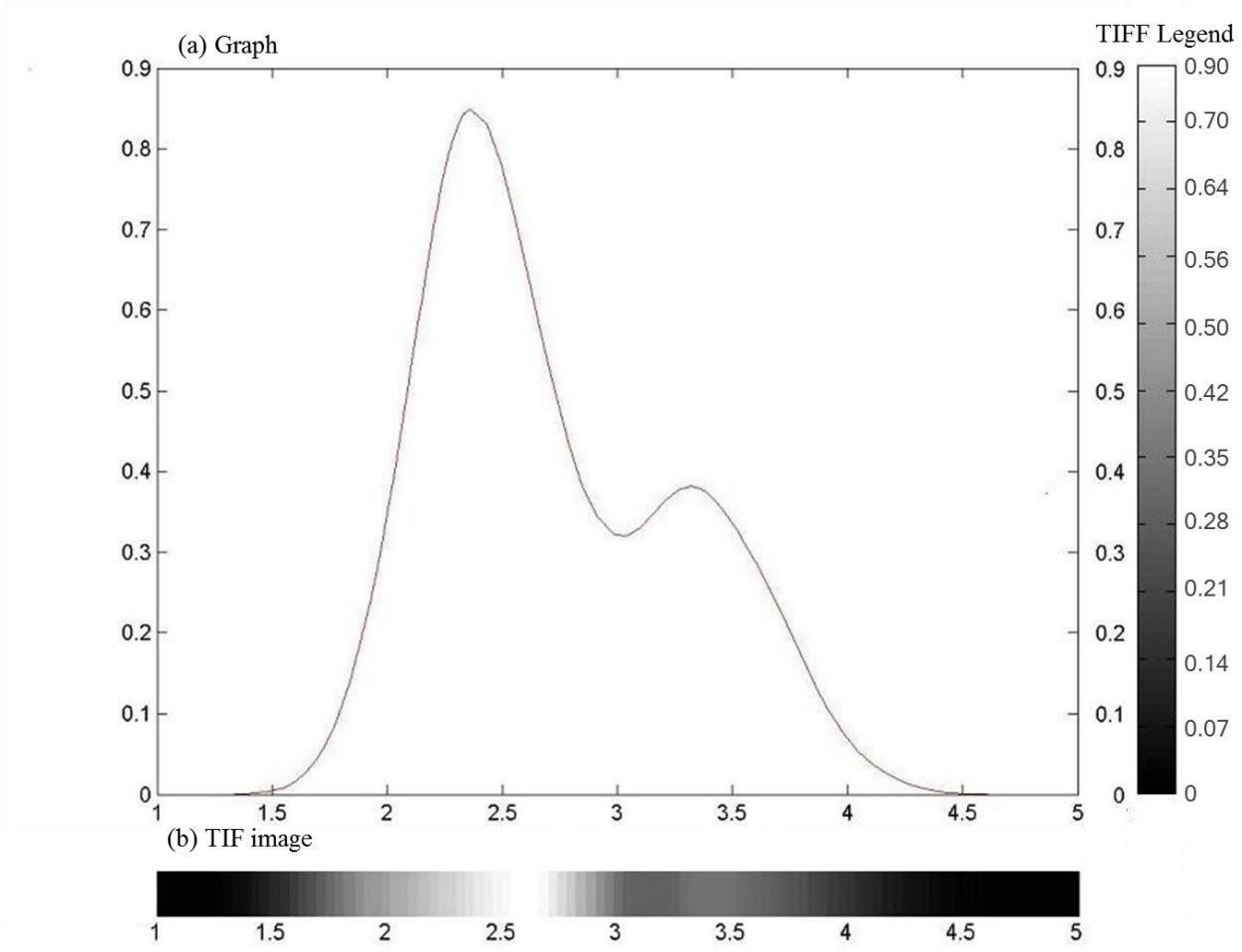

Figure 8.10: (a) Graph and (b) TIFF of the same curve. The x-axis is the same in both the graph and the TIFF image. In the graph, the height on the y-axis is density of the wing span of the aircraft. In the TIFF image the density of the wing span is expressed as the shade of the image instead of the height on the y-axis. The heights of the y-axis of the graph is from 0 to 0.9, this corresponds to the image colours from black to white. High values on the graph are shown as light on the image and low area on the graph are shown as black on the image.

8.2.5 Calculating the LDDMM mapping

One way to describe Large Deformation Diffeomorphisms is as a coordinate change between two images. The coordinate changes discussed in the previous sections map a curve from one grid or background space to another in order to prepare for the diffeomorphism. These different parameterisations are summarised in Table 8.1. Each of the grids represents a different background space and therefore a different expression of the same curve, just as the unit circle can be expressed as $x^2 + y^2 = 1$ in Cartesian coordinates or $r(\theta) = 1$ in polar coordinates. Through the steps of this method the curve has been expressed on a progression of different coordinate systems. To review, in step one of Table 8.1, each image was estimated on a fixed, regular Cartesian system of coordinates during kernel smoothing. In step two of Table 8.1, the images were moved to an unstructured surface grid with surface parameterisation. The coordinate system in step three was the unstructured surface grid relative to the reference surface created with landmarking; this coordinate change was made to facilitate image comparison because it creates a one-to-one matching of like areas on the surfaces. In step three of Table 8.1, the image grid is used and values are restricted to integers from 0 to 255. Throughout this process, each image has its own grid, but there is also a natural mapping between image grids.

Step four of Table 8.1 is the step of most interest. The deformation is a smooth curve that changes an image from its grid to another image's grid. The mapping or coordinated change via LDDMM is done and then the change is measured with the prescribed metric. The definition of a diffeomorphic coordinate change is an isomorphic mapping that operates on a smooth manifold, $I_0 \circ \alpha^{-1} = I_1$. By the definition of an isomorphism, the inverse of the mapping must exist. As well as a coordinate change, the LDDMM calculation is also expressed in terms of a curve, ϕ_t, defined for $t \in [0,1]$ and $\phi_t = v(\phi_t)$, $t \in [0,1]$ with the starting point defined as the identity matrix, $\phi_0 = id$ and whose endpoint, ϕ_1, is the inverse of the diffeomorphic coordinate change,

$\alpha = \phi_1$. Figure 8.12 shows an example mapping with interim curves.

Many curves meet this criterion, but it the curve of the shortest length which is of interest. Because the background space is not a fixed regular Cartesian mesh and the mapping curve is restrained to be perpendicular to the surface as shown in Figure 8.11, the distance along the curve is not a trivial measurement. The length of the mapping curve requires a minimization problem expressed as $argmin_{v:\phi_t = v_t(\phi_t)} \left(\int_0^1 \|v_t\|_v^2 \, dt \; + \left\| I_0 \circ \phi_t^{-1} - I_1 \right\|_{L_2}^2 \right)$. In this, the first term, $\|v_t\|$, measures the distance with a Sobolev norm and the second term is the error. Minimising the error term enforces matching of the two images. The result is a mapping curve that connects the two images, and is simultaneously the shortest path and closest match of the images. Looking at Figure 8.12, the error is the difference between the final mapping at timestep 20 and the target curve.

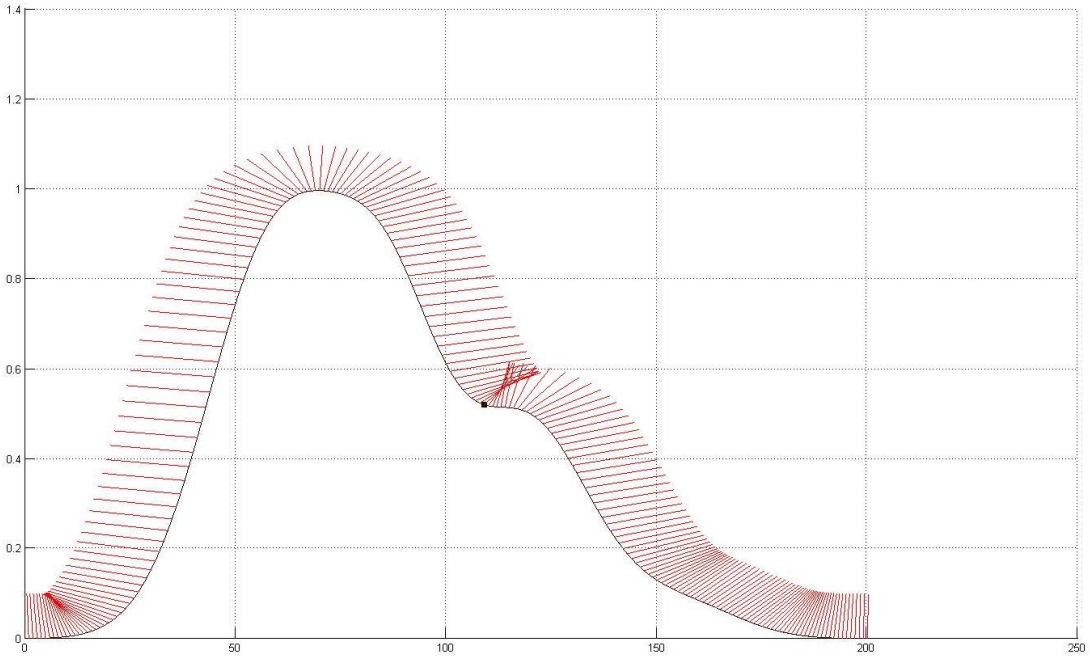

Figure 8.11: Normal vectors for a curve. The LMMDD algorithm restrains motion to the normal direction.

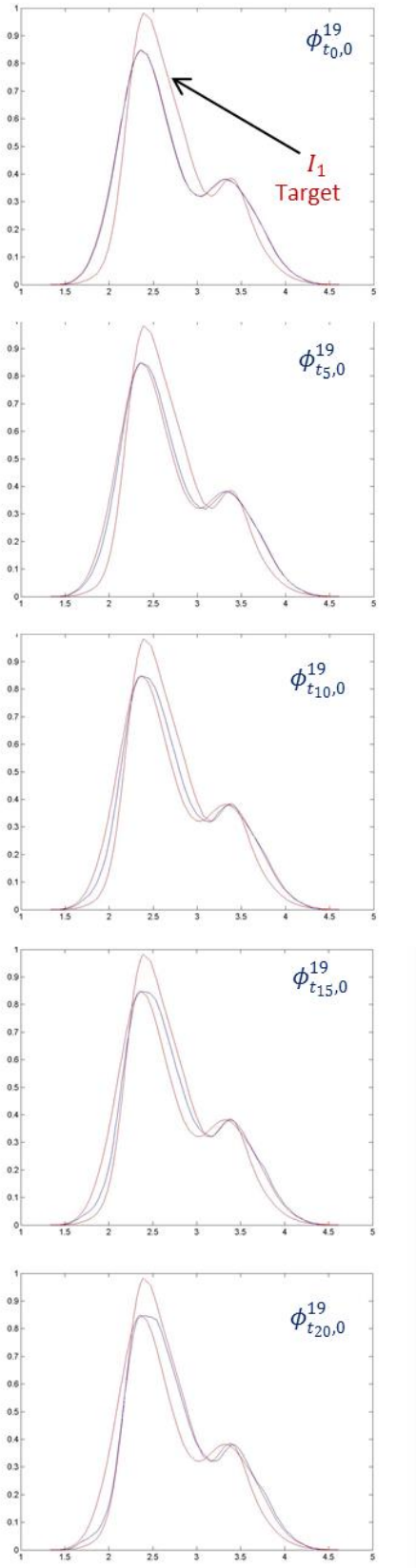

Figure 8.12: The source and target curves are shown in red and the interim mapping curve is shown blue. The interim mapping curve is shown at timesteps 0, 5, 10, 15 and 20. This is the 19th estimation of the curve produced by the 19th iteration of the algorithm.

The minimisation problem is performed at each point on the grid and the resulting unique mapping for each point is expressed as a vector field for each time point. The method uses the Euler–Lagrange equations for fluid flow. The estimate of the shortest path is performed following Beg *et al.*'s (2005) method with a modification to the stopping criteria. The estimate will be done iteratively, with the first estimate or iteration being the k = 0 estimate. For the first iteration, the starting image is assumed to be equal to the end image, which is expressed by setting the velocity for iteration $v_{t_j}^k = 0$ and the gradient of the velocity to $\nabla_{v^k} E_{t_j} = 0$. The vector field that estimates the curve is the identity matrix $\phi_{t_j}, T = Id \ \forall \ t_j \in [0, T]$. The maximum number of iterations is not known at the outset. The maximum number of estimates is set to *K* with each iteration being more accurate than the prior. For each estimate $k = 1, \ldots K$, the following steps are performed:

(1) The velocity at time *k* + 1 is estimated, $v^{k+1} = v^k - \epsilon \nabla_{v^k} E$, where E is an energy function that measures the goodness of fit of the transformation and ϵ is a small amout of noise.

(2) If needed, the velocity field is reparameterised so that the curve has a constant speed at each timestep.

(3) Using the equation $\phi_{t_j,T}^{k+1}(y) = \phi_{t_j,T}^{v}(y + \alpha)$, the *k* + 1 estimate of the mapping curve is estimated at each timestep, $j = N - 1$ to $j = 0$, for the backwards estimate of the curve based on the target surface, $\phi_{t_j,T}^{k+1}(y)$.

(4) Using the equation $\phi_{t_j,0}^{k+1}(y) = \phi_{t_j,0}^{v}(y - \alpha)$,The the *k* + 1 estimate of the mapping curve is estimated at each timestep, $j = 0$ to $j = N - 1$, for the forwards estimate of the curve based on the source surface, $\phi_{t_j,0}^{k+1}(y)$.

(5) The interim images associated with Step 3 are calculated for all timesteps, $j = N - 1$ to $j = 0$, showing the estimate of the diffeomorphism built from the target image. The image at each timestep is $J_{t_j}^1 = I_1 \circ \phi_{t_j,T}^{k+1}$.

(6) The interim images associated with Step 4 are calculated for all timesteps, $j = 0$ to $j = N - 1$, showing the estimate of the diffeomorphism built from the source image. The

image at each timestep is $J_{t_j}^0 = I_0 \circ \phi_{t_j,0}^{k+1}$.

(7) For each image in Step 6, the gradient is calculated, $DJ_{t_j}^0$.

(8) For all timesteps, $j = 0$ to $j = N - 1$, the Jacobian is calculated, $|D\phi_{t_j}|$.

(9) For all timesteps, $j = 0$ to $j = N - 1$, the gradient, $\nabla_{v^{k+1}} E$, of the velocity, v^{k+1} , is calculated using the equation $(\nabla_v E_t)_v = 2v_t - k\,(\frac{2}{\sigma^2}|\,D\phi_{t,1}^v|\,\nabla J_t^0(J_t^0 - J_t^1))$.

(10) Calculate the L_1 error term for iteration k + 1, $L_1 = \left|I_0 \circ \phi_{t_{N-1},0}^{k+1} - I_1\right|$.

(11) Stop if the L_1 error term for iteration $k + 1$ is equal to (or greater than) the L_1 for iteration k.

(12) Calculate the length of the mapping curve using the equation $\text{Length}\left(\text{Id}, \phi_T^{v^k}\right) = \sum_{j=1}^{N-1} \left\| v_{t_j}^k \right\|_v \delta t$, which is an estimate of the distance metric between the two surfaces.

The stopping criteria was changed slightly to a comparison of the mismatch error instead of the test of gradient change against a threshold used by Beg *et al. (2005)* in Step number 10 have been revised so the L_2 or absolute error is used instead.*(2005)*. Due to this change Step 10 has been revised to use the L_1 or absolute error. This error is not the value being minimised, but it will stop decreasing when the L_2 error that is in the joint minimisation stops decreasing.

Stopping criteria were changed from a test of the gradient change against a threshold used by Beg *et al. (2005)* to a comparison of the mismatch error. Either criterion should signal that the curve's estimation is no longer changing. Because a steepest descent scheme is in use, there is no guarantee that the global minimum has been reached (Arfken, 1985). The steepest descent is only certain of finding a local minimum and the choice of starting point will affect the search area. To ensure that a local minimum has been reached with the steepest descent algorithm and not just a plateau, the algorithm was run for three times longer than it took for the mismatch error to stop decreasing.

In this way, two curves are simultaneously estimated, one is an estimate of the mapping of I_0 to I_1. The endpoint of this mapping is the image $J_1 = I_0 \circ \phi$. The other is an estimate of the mapping of I_1 to I_0. The endpoint of this mapping is the image $J_0 = I_1 \circ \phi$. The LDDMM results obtained are consistent with Beg *et al.*'s implementation of LDDMM despite the slight change in stopping criteria. A diagram of the components of the two mappings and their relationship to the starting images is shown in Figure 8.13.

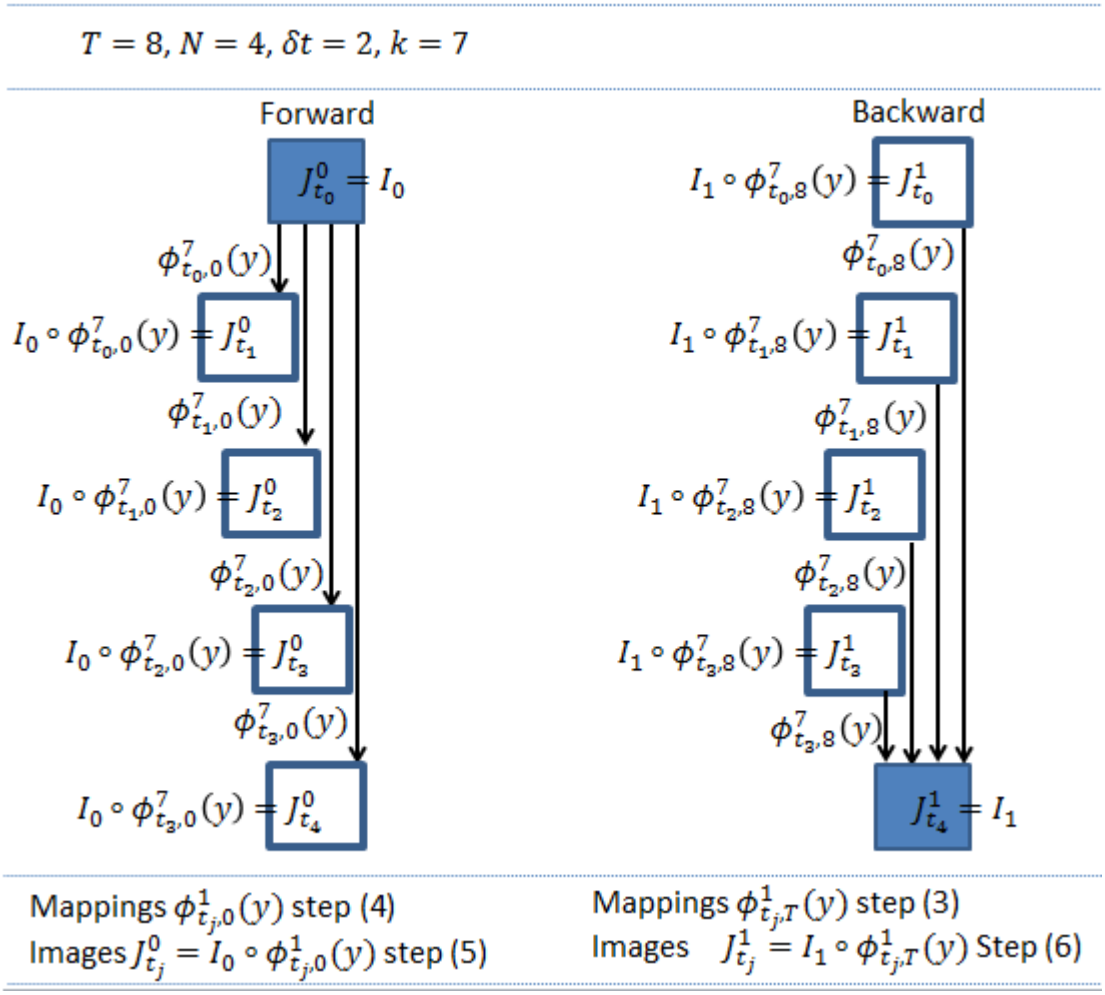

Figure 8.13: Diagram detailing the relationship between the source and target image, the interim curve estimates and mapping the functions.

8.2.6 The LDDMM distance metric

The LDDMM metric is a by-product of the mappings calculated in the previous section, and this mapping jointly minimises both the arc length and the mismatch error. The metric is simply the first term of the LDDMM minimisation equation, $\int_0^1 \|v_t\|_v \, dt$, which is a distance measure of the shortest smooth paths between the surfaces. To draw proper conclusions, this metric must be used in conjunction with the mismatch error. The mismatch error is a measure between the intended endpoints of the mapping (the target curve) and the actual endpoint of the mapping (J_1 or timestep 1) Because the LDDMM calculation is a joint minimisation on both the arc length and the mismatch, the mismatch term indicates whether an acceptable mapping has been performed. If the mismatch is too large, the endpoint of the mapping curve, J_1, may have little resemblance to the target curve. In the initial step of the LDDMM algorithm, the J_1 curve is equal to the J_0 curve. This means that the distance along the curve connecting J_0 and J_1 is zero and the mismatch term is very high. At each iteration of the algorithm, the length of the curve connecting J_0 and J_1 is increased and the mismatch error is decreased. When the mismatch stops decreasing, the best possible match has been found, but it is still necessary to check that the mismatch has decreased to an acceptable value and that the J_1 and I_1 curves are acceptably close.

There are several differences between this method and the method of comparing displacement (L_1) or the mean integrated squared error (MISE or L_2). Normally, L_1 and L_2 comparisons are done on the Euclidean grid and points are aligned by x-coordinates. This gives the minimum or straight line distance without regard to the features of the curve. The LDDMM method forces the paths to follow the norms of the surface and the landmarking process aligns the features of the curve. Because the curve is constrained to be smooth and the movement begins in the direction of the norms, the resulting curves are more indicative of actual movement patterns. If the curves

are the density of a population at two points in time, the mapping shows the forces which would deform one image into another with the least energy.

Table 8.2 shows a comparison of the LDDMM, the MISE and the displacement of the three surfaces in the example aircraft dataset. In the dataset each surface describes the distribution of wing spans in aircraft design during a time period. Two values are shown for MISE and displacement. The first is the value for the landmarked surfaces where the features of the curves are aligned. The second is the MISE and the displacement for the original surfaces before landmarking where the surfaces are aligned by *x*-coordinates. Finally, the table shows the LDDMM metric and error. The LDDMM compares the landmarked surfaces.

Comparison		LDDMM			Landmarked		Original surfaces	
Start	End	Iterations	Metric	Error	MISE	Displacement	MISE	Displacement
3	1	19	0.00336	0.08876	6.2109	0.09967	5.6233	0.08659
1	3	19	0.00219	0.09472	6.2109	0.09967	5.6235	0.08659
1	2	19	0.00254	0.09684	6.2305	0.10130	3.0843	0.07453
2	1	19	0.00180	0.09224	6.2305	0.10130	3.0843	0.07453
2	3	19	0.00008	0.03379	0.6773	0.03804	2.2133	0.06744
3	2	19	0.00009	0.03530	0.6773	0.03804	2.2133	0.06744

Table 8.2: LDDMM, MISE and displacement metrics for the aircraft dataset. MISE and displacement are shown for both the original surface and the landmarked surface.

The three measurements (LDMM, MISE, and displacement) are in agreement that the most similar surfaces are Surfaces 2 and 3. The smallest value in the displacement of original surfaces column is, 0.06744, the comparison of Surfaces 2 and 3. Similarly, the smallest values in the LDDMM Metric, LDDMM Error, Landmarked MISE, Landmarked Displacement, Original

MISE and Original Displacement are 0.00008, 0.03530, 0.67733, 0.03804 and 2.21333 respectively. The difference between the LDDMM beginning at Surface 2 and ending at Surface 3 (0.00008, 0.03379) has a negligible difference to the partner metric beginning at Surface 3 and ending at Surface 2 (0.00009, 03530).

The distance between Surfaces 1 and 3 and that between Surfaces 1 and 2 are much more interesting, the metrics for the original surfaces (MISE and the displacement) suggest that Surface 1 is closer to Surfaces 2 than to Surface 3 (a MISE of 3.08431 *vs*. a MISE of 5.623 and a displacement of 0.07453 *vs.* 0.08659), but the metric for the landmarked surfaces (displacement and MISE) are nearly identical. The distances between the landmarked surfaces have reversed from those of the original surface comparisons. The displacement from Surface 1 to Surface 3 is 0.09967 for the landmarked surface, this is less than the displacement from Surface 1 to Surface 2, 0.10130. The MISE is also reversed, for the landmarked surfaces it is 6.21098 for the Surface 1 to 3 comparison and 6.23059 for the Surface 1 to 2 comparison. The LDDMM metrics for Surface 1 to 3 overlap with those for the change from Surface 1 to 2; there is no clear difference between the distances.

Comparison		Iteration 19		Iteration 1		Iteration 1/ Iteration 19
Start	End	Metric	Error	Metric	Error	
3	1	0.00336	0.08876	0	0.09967	0.8905
1	3	0.00219	0.09472	0	0.09967	0.9503
1	2	0.00254	0.09684	0	0.1013	0.9559
2	1	0.0018	0.09224	0	0.1013	0.91056
2	3	0.00008	0.03379	0	0.03804	0.88827
3	2	0.00009	0.0353	0	0.03804	0.9279

Table 8.3: Comparison of first and last iteration of LDDMM for aircraft dataset.

The iterations of the LDDMM metric Table 8.3 exemplify how the LDDMM metric and error measurements are made. The algorithm starts off at Iteration 1 with the error being equal to the displacement of the landmarked surfaces. The metric is then increased as the error decreases. The stopping criterion is met when there is no reduction in error for successive iterations. The LDDMM calculation can be thought of as breaking the displacement metrics into two parts: the LDDMM metric and the error. The metric is the amount of the difference that can be accounted for with a deformation. The error is the residual distance that cannot be smoothly deformed.

Curves with similar shape can be deformed with little residual error. Looking at the ratio between Iteration 1's error or displacement and Iteration 19's error, very little decrease is seen in the error. This means that very little of the difference can be accounted for by shape similarity.

Returning to the sail boat analogy discussed in the introduction to this chapter (Section 8.1) which compares the LDDMM mapping to measuring the distance between two shores by sailing a fleet of small boats in tight formation and using the average acceleration of the boats as a measure of distance, several observations can be made. The LDDMM movement model is constrained to be smooth and in the direction of the surface norm. If a boat sails from a known starting point, the boat starts facing a known direction and the boat must sail to a finishing point and arrive facing a given direction. The boat has to sail smoothly into turns and sometimes it is not possible to arrive exactly at the target in one smooth path. The shortest displacement path also ignores the direction of the boat at the beginning and the end of the journey. The norm of the surface represents the orientation of points on the surface or the boats on the starting shore.

8.3 Case Studies

8.3.1 Starkey

The Starkey dataset consists of two models of the bias in Automated Telemetry System (ATS) data collected at the Starkey Experimental Forest and Range in Oregon, USA. In the study elk (*Cervus elaphus*), mule deer (*Odocoileus hemionus*), and cattle were fitted with radio collars and their location was tracked over a 10,102-ha reserve surrounded by a game-proof fence. The bias models each give the probability a fix at a location will be successful. The locations are defined to be 30-by30 metre grids squares. The MCMC model developed in Chapter 8 is referred to as model 1 and the Johnson *et al.* (1998) model is referred to as model 2. For more detail on the models see Chapter 8, for more detailed information on the ATS, the tracking data, and habitat information see the U.S. Forest Service web site (http://www.fs.fed.us/pnw/starkey/).

LDDMM calculation for the dataset begins with conversion to image parameter space, because the smooth surfaces have already been estimated, landmarked, and discretised. For this dataset the latitude and longitude are used to landmark the surfaces. The models of bias in ATS are each smooth three dimensional surfaces, two dimensions are the latitude and longitude and the third is a weight used for fix rate bias correction. Each surface gives the probability of a fix succeeding for each location in the study site. The surface is discretised and stored in a two dimensional matrix where each cell in the matrix represented a 30-by-30 metre grid square and the value of the cell is the bias correction weighting estimated by the model. The study site is not square, so zeroes were placed in the parts of the matrix outside the study area. Instead of using maxima and minima to align the surfaces, latitude and longitude were used.

Comparison		LDDMM			MISE	Displacement
Start	End	Iterations	Metric	Error	Metric	Metric
1	2	3	0.4644	1.5027	384.67	1.5085
2	1	11	0.2316	1.5070	384.67	1.5085

Table 8.4: Comparison of LMMDD, MISE, and displacement metrics for Starkey dataset.

The matrices were converted to image parameter space by creating a pixel, (p_r, p_b, p_g), for each cell in the matrix with the formula $p_r = p_b = p_g = \left\lfloor \frac{w - w_{min}}{w_{max} - w_{min}} * 256 \right\rfloor$, where w is the weight in the cell, w_{min} is the minimum weight in all cells of all matrices, and w_{max} is the maximum weight of all cells in all matrices models. The result is two grey-scale TIFF images which are ready for the LDDMM calculation. The results of the LDDMM calculations are shown in Table 8.4 and Table 8.5. All results in the tables are shown as weights, not the image scale used in the TIFF.

Comparison		Final iteration of LDDMM			Iteration 1of LDDMM	
Start	End	Iteration	Metric	Error	Metric	Error
1	2	3	0.4644	1.5027	0	1.5085
2	1	11	0.2316	1.5070	0	1.5085

Table 8.5: Comparison of the first and last iterations of the LDDMM metric for the Starkey dataset.

The result of the LDDMM calculation shows very little reduction in the displacement measured between the two matrices before the LDDMM mapping. As discussed earlier, the LDDMM metric and the LDDMM error break the displacement into two components. The metric is the amount of the difference that can be accounted for with a deformation and the error is the residual distance that cannot be smoothly deformed. The small reduction in displacement means that a very small amount of the difference between the surfaces is due to shape similarity. The likely explanations for this are that the second model is measured on a coarser grid than the first

and has a distinctly different shape.

The results for comparing the bias correction weights for the Starkey dataset using LDDMM confirm that the smooth surface estimated for bias correction weight in the prior chapter is distinctly different from the shape of the smooth surface estimated by Johnson *et al.* (1998). The similarity of the LDDMM weight to the displacement may indicate that different landmarking would give a better result. Additional comparisons could be made by landmarking of the maxima and minima before performing the LDDMM or using calculation model 1 on the same coarser grid as model 2.

8.3.2 Ozone

The ozone dataset (Cleveland *et al.*, 1992) measures air quality conditions in New York City. The data consist of measurements for ozone in parts per million, solar radiation, temperature in degrees Fahrenheit and wind speed in miles per hour. The data were recorded once per day over the period of the study. The dataset is used as the example dataset by Bowman (2007) in his paper on the comparison of nonparametric surfaces. Bowman (2007) separated the radiation into high and low groups, shown as o and + respectively in the plots. The lack of data separation is made clear in both temperature vs. ozone (Figure 8.15) and wind vs. ozone plots (Figure 8.15).

Bowman compares five smooth surfaces produced by three models. The same five surfaces were compared with LDDMM. The LDDMM results are compared with Bowman's results and new insight provided by the LDDMM metric are discussed. Bowman performs hypothesis testing with approximate degrees of freedom and is able to test hypotheses. This thesis will cover only the utility of the LDDMM metric and whether it confirms the results of Bowman's work. As currently presented, formal statistical inference and hypothesis testing cannot be conduct with

the LDDMM metric. The additional steps needed to perform inference with LDDMM are discussed in Section 8.1.2.

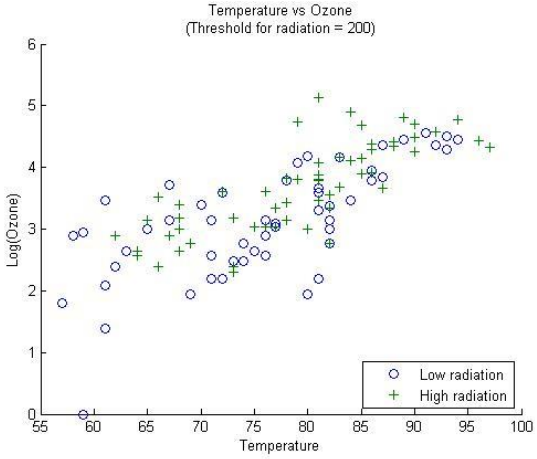

Figure 8.14: Temperature vs. ozone for the ozone dataset.

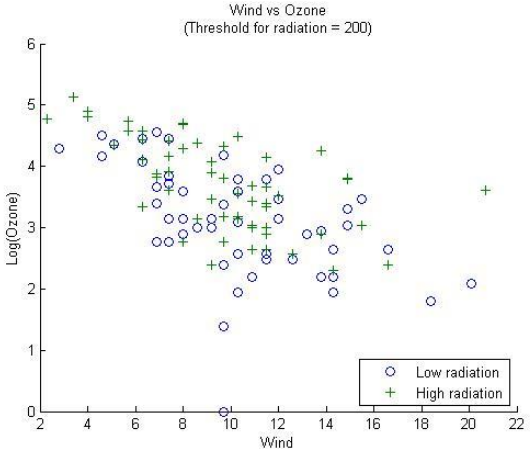

Figure 8.15: Wind vs. ozone for the ozone dataset.

The three models considered by Bowman are a single surface, two independent surfaces and two parallel surfaces. The single surface model uses all *n* observations (x_{1j}, x_{2j}, y_j), $j \in 1, \ldots, n$, where the variable $x_1$is wind, x_2 is temperature and y_j is ozone. The equation for model is

$$y_j = m_j(x_{1\mathrm{j}}, x_{2\mathrm{j}}) + \epsilon_j\,, \qquad [8.10]$$

where ϵ_j is the error. This surface acts as the null hypothesis for the other models during hypothesis testing. A graph of this surface is presented in Figure 8.16 (e).

The two independent surfaces model has the equation

$$y_{ij} = m(x_{1ij}, x_{2ij}) + \epsilon_{ij}, \qquad [8.11]$$

for groups $i \in 1{,}2$. Where the observations are $j \in 1 \ldots n_1, 1..n_2$, $n_1 + n_2 = n$ and the ϵ_{ij} is the error term. Observations in group one, $i = 1$, have solar radiation ≤200 and observations in group two, $i = 2$, have solar radiation >200. Graphs of these surfaces are presented in Figure 8.16 (a) and (b).

The two parallel surfaces model uses an additive parameter to separate parallel surfaces. The equation for this mode

l is

$$y_{ij} = \alpha_i + m(x_{1ij}, x_{2ij}) + \epsilon_{ij}, \qquad [8.12]$$

where α_i is the additive parameter, $i \in 1{,}2$ are the groups, $j \in 1 \ldots n_1, 1..n_2$, $n_1 + n_2 = n$ are the observations, and ϵ_{ij} is the error term. Figure 8.16 (c) and (d) contain graphs of these surfaces. Just as in the two-surface model, observations in group one, $i = 1$, have ≤200 solar radiation. A single surface $m(x_{1ij}, x_{2ij})$ is fitted from all data, but the additive parameter α_i creates a vertical shift between the two groups of data. The additive parameter is set to zero for the high radiation group, $\alpha_1 = 0$. The single-surface model can be stated as a parallel surface model with $\alpha_2 = 0$.

The surface fitting was done in R (R Core Team, 2012) using the '`sm`' package (Bowman and Azzalini, 2010). The fitting is done via nonparametric regression with a Gaussian kernel and bandwidth, $h = (h_1, h_2)$. The bandwidth is calculated using the Corrected Akaike Information Criterion (AIC_c) proposed by Hurvich *et al.* (1998). The bandwidth minimises $\log(\hat{\sigma}^2) + 1 + (2tr\{S\} + 1/ (n - tr\{S\} - 2)$, where S is the smoothing matrix for the surface estimate $\hat{m}$, $\hat{m} = Sy$. Additionally, h is constrained to be $(h_0\sigma_1, h_0\sigma_2)$ so that the same amount of smoothness

is applied to both the wind and temperature dimensions. This results in a regression surface defined as $\widehat{m} = (x_1, x_2) = \hat{\alpha}$. The surface is defined for all points (x_1, x_2) with the minimising equation

$$\min_{\alpha,\beta,\gamma} \sum_{i=1}^{n} \{y_i - \alpha - \beta(x_{1i} - x_1) - \gamma(x_{2i} - x_2)\}^2 w(x_{1i} - x_1; h) w\,(x_{2i} - x_2; h_2). \quad [8.13]$$

To mirror Bowman's result, the following bandwidths were used: $h_1 = 2.7$ for wind and $h_2 = 7.1$ for temperature, derived from the single-surface as the bandwidth for all surfaces.

A total of five surfaces are fitted. The first surface is fitted with the single surface model using all data. The surface is referred to as the all-data surface. The second and third surfaces are fitted with the two independent surfaces model and referred to as the high-radiation surface and the low-radiation surface. The final two surfaces are fitted simultaneously with the two parallel surfaces model and are referred to as the high- and low-radiation parallel surfaces.

Now that the smooth surfaces have been calculated, the next steps for comparing the models using LDDMM are identifying the landmarks, surface discretisation and conversion of the surfaces to images. Because all of the surfaces are fitted from the same observations, the *x*-values of the wind and temperature will act as the landmarks. Due to this, landmarking to align maxima and minima will not be performed. The surface discretisation is performed using a 250 × 250 point grid and equally spaced grid points along the two *x*-axes. The conversion to the image parameter space follows the method described earlier.

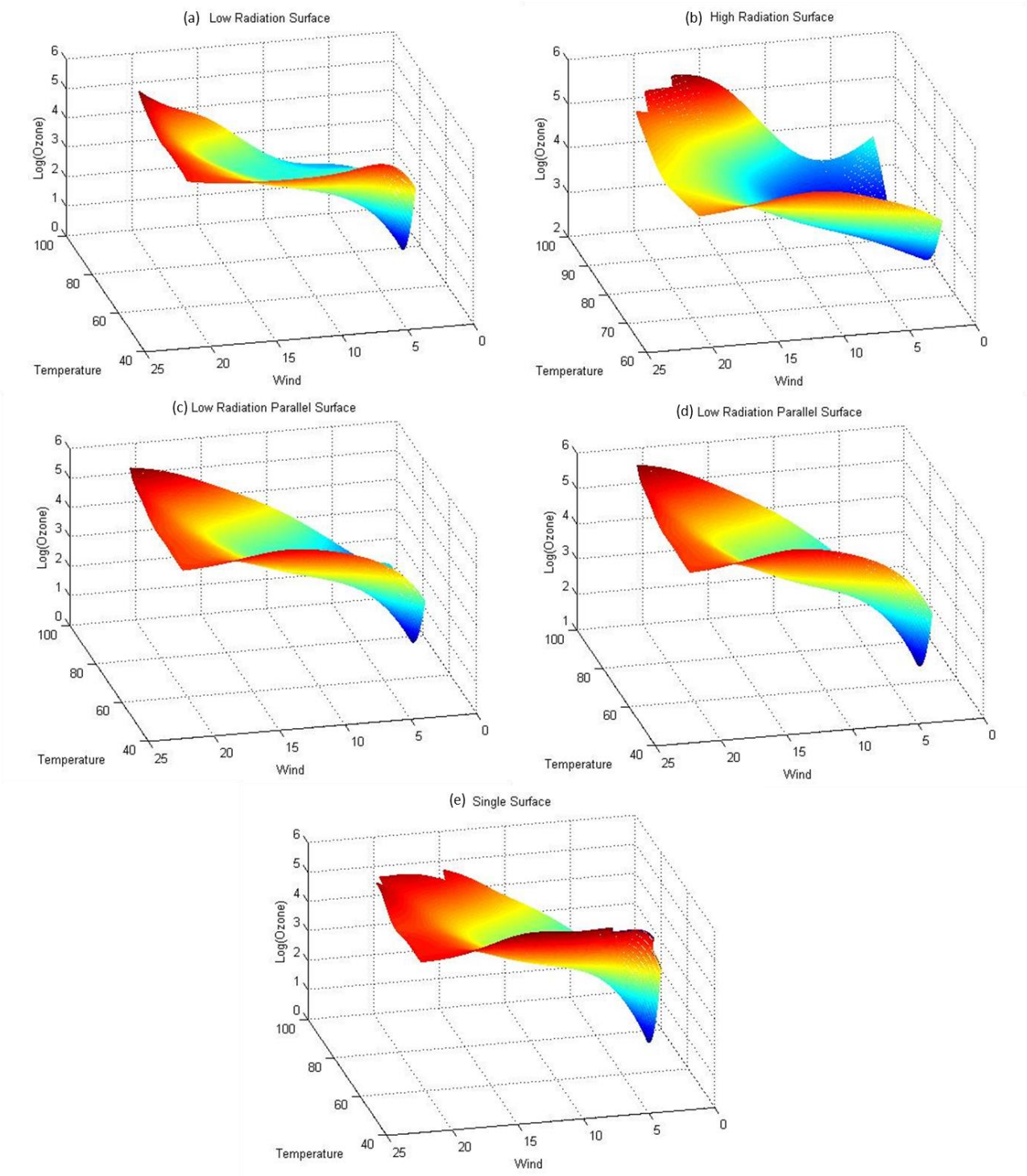

Figure 8.16: Five surfaces fitted to the ozone data using different models. Surfaces (a) and (b) are fitted with the high- and low-radiation models respectively, (c) and (d) are fitted with the high- and low-radiation parallel models respectively, and (e) is fitted with the single-surface model.

(a) LDDMM metric		Target model				
		Low-radiation	High-radiation	Low-radiation parallel	High-radiation parallel	Single-surface
Source model	Low-radiation	0	3.9810	0.4134	0.4134	2.4313
	High-radiation	5.1715	0	5.6959	5.6959	13.7814
	Low-radiation parallel	0.5149	4.4372	0	0	2.1948
	High-radiation parallel	0.5149	4.4372	0	0	2.1948
	Single-surface	2.3512	10.4622	2.1959	2.1959	0

(b) LDDMM error		Target model				
		Low-radiation	High-radiation	Low-radiation parallel	High-radiation parallel	Single-surface
Source model	Low-radiation	0	0.1361	0.0363	0.0363	0.0849
	High-radiation	0.1317	0	0.1341	0.1341	0.1964
	Low-radiation parallel	0.0359	0.1403	0	0	0.0714
	High-radiation parallel	0.0359	0.1403	0	0	0.0714
	Single-surface	0.0840	0.2053	0.07010	0.0701	0

(c) Number of iterations		Target model				
		Low-radiation	High-radiation	Low-radiation parallel	High-radiation parallel	Single-surface
Source model	Low-radiation	1	959	705	705	1999
	High-radiation	1999	1	723	723	1999
	Low-radiation parallel	882	828	1	1	1999
	High-radiation parallel	882	828	1	1	1999
	Single-surface	1999	1999	1999	1999	1

Table 8.6: LDDMM results of the ozone dataset. (a) LDDMM metric, (b) LDDMM error and (c) the number of iterations required for fitting.

The LDDMM metric can now be calculated, selecting two LMMDD metrics for each pair of surfaces. There are five surfaces, so a total of 25 comparisons were made. The results are shown in Table 8.6, along with the corresponding errors and the number of iterations preformed.

Looking at the LDDMM metric and the error, there are some interesting observations that confirm the models used to create them. When each surface is compared to itself, the resulting difference is zero and it only takes one iteration to reach this result, just as expected. The shape difference between the high and low parallel surfaces is zero. This is expected because the shapes are identical, although they are shifted on the ozone axis by the α value. The smallest non-zero LDDMM metric is 0.413 (between the parallel surfaces and the two parallel surfaces). This suggests that the shapes of these three surfaces are the most alike (excepting of course the identical surfaces). The largest LDDMM metric between two surfaces is 13.7814, which is the difference between the high-radiation model and the single-surface model. This means that the shape of the high-radiation model and the single-surface model are more different than any other pair of surfaces. Note that several of the mappings ran the maximum number of iterations (1999) without meeting the stopping rule. All of these estimates were later run for 9999 iterations and the results were only slightly improved (data not shown). Changing the stopping criteria to use a threshold for improvement similar to the methodology of Beg et al. (2005) would curtail these long-running calculations that do not improve the result.

Bowman (2007) makes two hypothesis tests. In the first, the null hypothesis is a single surface, and the alternative hypothesis is two surfaces. Bowman's first observation is that there is strong evidence for the two-surface hypothesis. This is confirmed by the high LDDMM values between the high-radiation surface and the single surface (13.7814 and 10.462245), as well as the high LDDMM values between the high-radiation surface and the low-radiation surface (5.171583 and

3.9810168). The high-radiation surface is different from both the low-radiation surface and the single surface that acts as the null hypothesis.

(a) L_1		Target model				
		Low-radiation	High-radiation	Low-radiation parallel	High-radiation parallel	Single-surface
Source model	Low-radiation	0	0.1242	0.0899	0.0454	0.0725
	High-radiation	0.1242	0	0.0482	0.0975	0.0595
	Low-radiation parallel	0.0899	0.0482	0	0.0605	0.0327
	High-radiation parallel	0.04546	0.0975	0.06054	0	0.0444
	Single-surface	0.07255	0.05957	0.0327	0.0444	0

(b) MISE		Target				
		Low-radiation	High-radiation	Low-radiation parallel	High-radiation parallel	Single-surface
Source model	Low-radiation	0	5.0342	2.6313	0.8688	2.0221
	High-radiation	5.0342	0	1.3586	3.4323	1.8184
	Low-radiation parallel	2.6313	1.3586	0	0.9353	0.5188
	High-radiation parallel	0.8688	3.4323	0.9353	0	0.9259
	Single-surface	2.0221	1.8184	0.5188	0.9259	0

Table 8.7: (a) $L_1$and (b) MISE for the ozone dataset surfaces.

The second comparison Bowman makes is between the single-surface null hypothesis and the alternative hypothesis of two parallel surfaces. Bowman did not find enough evidence to reject the null hypothesis. This result is confirmed by LMMDD because the LDDMM values between the two parallel surfaces and the null hypothesis are small (2.196 and 2.196). In addition, the LDDMM between the parallel and nonparallel surfaces are larger than those between the parallel

surfaces and the single surface.

L_1		Target				
		Low-radiation model	High-radiation model	Low-radiation parallel model	High-radiation parallel model	Single-surface model
Source	Low-radiation model	0	0.1242191	0.089918525	0.0454653	0.07255
	High-radiation model	0.1242191	0	0.048265608	0.09759444	0.05957
	Low-radiation parallel model	0.0899185	0.0482656	0	0.06054313	0.03275
	High-radiation parallel model	0.0454653	0.0975944	0.060543129	0	0.04441
	Single-surface model	0.0725541	0.059571	0.032751375	0.04441014	0

MISE		Target				
		Low-radiation model	High-radiation model	Low-radiation parallel model	High-radiation parallel model	Single-surface model
Source	Low-radiation model	0	5.0342299	2.631328346	0.8688026	2.02214
	High-radiation model	5.0342299	0	1.358601464	3.43238969	1.81849
	Low-radiation parallel model	2.6313283	1.3586015	0	0.93532334	0.51886
	High-radiation parallel model	0.8688026	3.4323897	0.935323339	0	0.92598
	Single-surface model	2.0221378	1.8184901	0.518856173	0.92598162	0

Table 8.8: $L_1$and MISE for the ozone surfaces.

In conclusion, while a formal hypothesis test has not been conducted, the results of the LDDMM metric are consistent with the comparisons conducted by Bowman and appear to give a measure of the shape difference rather than the Euclidean distance between the surfaces. The results presented here are not directly comparable to Bowman's because LDDMM compares the fitted surfaces and Bowman uses the residual between the surfaces and the observations. This is shown by comparing the L_1distances and the MISE for the surfaces (Table 8.7). These traditional surfaces show a difference between the parallel surfaces, while the LMMDD metric shows no difference.

8.4 Discussion

One drawback of LDDMM is that for any two images, I_p and I_q , there two mappings and thus two metrics: the metric resulting from $I_0 = I_p$ and $I_1 = I_p$, and the metric resulting from $I_0 = I_q$ and $I_1 = I_p$. These metrics are not equal. This is because one set of metrics is constrained to move along the norms dictated by I_p and the other is constrained to the norms dictated by I_q. A mixture of these two mappings can be used. For each time point, t,

$$\frac{t}{T}\ vector_1 * \left(1 - \frac{t}{T}\right) vector_2. \qquad [8.14]$$

The combined mapping gives a smooth transition between the two surfaces and begins with motion in the direction of the source surface norm, and ends with motion in the direction of the target surface norm. One notable problem is that with this combined mapping, the two input mappings are minimised but there is no assurance that the combined mapping is minimised. Further research is needed into the suitability, the properties of the mapping and a suitable metric.

Another drawback to the metric is that there is no absolute scale. This means that if M_{12} is the metric from comparing images I_1 and I_2, and M_{34} is the metric from comparing two unrelated images I_3 and I_4, no conclusion can be drawn by comparing M_{12} and M_{34}.

The question about what the LDDMM metric actually measures arises naturally at this point. The Euclidian measures of shape, variance, skew, kurtosis, etc., are defined in Euclidian space and can be used on absolute grids such as the regular Cartesian mesh. The triangle inequality and other familiar properties of norms and distances hold. These standard measures of shape differences are not suitable for LDDMM for several reasons. One reason is that the mesh is not absolute, but is defined relative to the landmarking reference surface. This means that absolute distance is not defined. Another reason is that the distance of interest is the distance measured along the surface or the arc length of the curves joining the images, not the displacement between the start and endpoint. A more general measure of shape difference is required, and a more generalised space than Euclidean space must be used. The LDDMM metric operates in a Sobolev space and uses Sobolev norms to measure shape difference. Just as there are different measures of shape difference in Euclidean space, there are corresponding measures of shape difference in Sobolev space. There is a norm corresponding to each of the Euclidean shape measurements. LDDMM uses the L_2 norm, which is roughly equivalent to variance in Euclidean space.

The drawback of the LDDMM metric and all other metrics defined on a Sobolev space is that the triangle inequality does not hold, and thus they are not true distances. This is why the metric between I_0 and I_1 can be different from the metric between I_1 and $I_{0,}$ and the metric between I_0 and I_1 plus the metric between I_1 and I_2 can be shorter than the metric between I_0 and I_2. Because of the goal of measuring the distance along the surface and along the path of particle

displacement, the less restrictive Sobolev space is the correct choice, but with this choice comes the need for non-Euclidean metrics of shape difference.

8.5 Conclusion

The deformations originally developed to compare anatomical features are applied for the first time outside of computational anatomy to compare surfaces resulting from statistical models. One advantage of this shape comparison over standard shape comparisons is the process of landmarking that aligns like areas of the statistical models before comparing them. The second advantage is that diffeomorphisms measures the shape change along the curves of fluid flow so that the orientation of the points on the surface are accounted for as well as the displacement. The drawbacks are the lack of absolute scale, the loss of the triangle inequality, the complexity of the calculation, and the input required by a subject matter expert to perform the landmarking.

As a diffeomorphic shape comparison method, LDDMM shows many very desirable properties. It preserves surface features over many timesteps, whereas many earlier propagation techniques distorted features or lost the shape of the surface over time. It allows long timesteps by using semi-Lagrangian methods, thus speeding computation time. It preserves neighbourhood properties among points of the surface so that points that are close in the original surface are close in the transformed surface while points that are distant in the original surface are distant in the transformed surface. Small deformation techniques sometimes allow the surface to fold back on itself, e.g. turning a straight line into a figure eight, in doing so the distant points became close.

LDDM has been selected for this research because it allows a different transformation for each point. The LDDMM transformation allows each point to transform, but the transformations are

not independent. Each transformation is constrained by the neighbouring transformations. Bowman's parallel comparison is effectively a single transformation applied to one group of data. Bowman performs a simple vertical transformation that allows the comparison of two parallel surfaces. The LDDMM shift only acts in small areas of the surface, for example a shift in the density of animals in only one small area of the study region due to a shifting tree line or a change in land use. The parallel shift in Bowman's model must be vertical, LDDMM allows either vertical or horizontal shift. The deformations are applied via a series of composition so that a backwards mapping always exists. The resulting overall translation can be thought of as a curve.

LDDMM metrics can be used to test assumption of global similarity, the source and target surfaces are similar on a global level, and each point on the source surface is in the same neighbourhood in the target surface. This method can be applied to data collected without a sampling design which dictates the sampling location, such as GPS data, or administrative data collected for other purposes, such as the number of deer culled in a hunting season. If the data are collected over multiple years in a large study area and each year has good spatial cover, then a reasonable estimate can be made of the shape of the surface over the entire study area, but the same sites are not visited each year and the study area is likely to have changed during the period. The sites visited repeatedly can be used as landmarks for the surfaces. The hypothesis is that that two years show the same results. For the sites only visited in one year, a close site in the other years will be used for comparison.

Another case where this method could be used is in a comparison of difference in models derived from the same data such as the bias correction weight models in Chapter 8. One model is fitted with a subset of the data used to fit the second model, but both models are for the same

study site. The alignment of the surfaces is obvious because data were collected at the same study area, so physical location can be used for landmarking. Other observations from the study area that can be expressed as smooth surfaces can be compared with LDDMM. For example, a radio reception model, or an elevation map for the Starkey study area could be compared to the bias correction model.

It would be desirable to do statistical inference with these metrics, but further research is required before inference is possible. Following the framework of computational anatomy used by Grenander and Miller (1998) there are three areas of research to complete: (1) the computation of the deformation maps or metrics, (2) the computation of empirical probability laws, and (3) computational inference, testing and classification (Grenander and Miller, 1998). This chapter discusses only the computation of the deformation metrics for statistical surfaces.

8.5.1 Future work

Future work in this area would include performing formal statistical inference with LDDMM. This chapter is concerned only with mapping and how to apply the mappings developed for anatomical surfaces to regression surfaces. In order to do inference in computational anatomy, the mapping metric must be defined first. Then the mapping is carried out on a large population and the results are used to form a collection and define the variation of the population so that the empirical probability laws can be defined. Finally, these laws are used to make inferences via Bayesian classification.

9 Detecting anomalous regression surfaces with Diffeomorphic Demons

9.1 Introduction

In Chapter 8 the LDDMM algorithm was applied to compare statistical distributions in the context of environmental science. In this chapter, another diffeomorphic algorithm from computational anatomy, Diffeomorphic Demons (DD), is applied in a different context. The new context is finance, specifically the comparison of implied volatility surfaces. Instead of a metric to quantify shape differences as in the last chapter, the aim is detecting anomalous surfaces. From a large set of implied volatility surfaces, those surfaces that have a distinctly different shape will be selected.

As discussed in detail in Chapter 8, surfaces and shape comparisons are integral parts of medical imagery for analysing digitised images of the human anatomy. Anatomical surfaces such as the membranes around organs are now being mapped by technologies like Magnetic Resonance Imaging (MRI), Diffusion Tensor Magnetic Resonance Imaging (DTMRI) and Functional Magnetic Resonance Imaging (fMRI) (Beg *et al.*, 2005). To use of these scans as diagnostic tools, medical professionals to locate regions of interest and compare them, using expert knowledge, to the expected size and shape to detect anomalies.

Morphometric algorithms automate the process of comparing and contrasting images. Typically,

individual scans are compared with a template that is a composite of multiple scans (Miller *et al.*, 2002). This is not a straightforward task and simple point-wise comparisons yield uninformative results. Initial deformation of images is done to align the image to the template while preserving neighbourhood properties (Cachier *et al.*, 2003) before the comparison is performed. In Chapter 8 the landmarking method of alignment was discussed.

Both LDDMM and DD have two components: a distance of mapping component and a mismatch component. Both processes involve minimizing both of these components simultaneously. In Chapter 8 the mapping term was used as the basis for the resulting metric because it represented the distance between the shapes. Because this chapter is interested in anomaly detection, the metric will be based on the mismatch term, and the mapping term will be viewed as a coordinate change that aligns a particular day's trading to the reference surface.

The DD algorithm is comparatively easier to calculate than LDDMM (Hernandez *et al.*, 2008). In medical contexts, it has been shown to have similar results to the LDDMM algorithm (Hernandez *et al.*, 2008). The DD algorithm is performed here without any landmarking and is therefore a more automated process than that developed in Chapter 8. This allows the rapid comparison of a large number of surfaces. The intuitive similarity decisions made by the human brain is automated to produce a metric that measures the similarity of the shape of two images. This metric is used on a series of images to find those that are anomalous. The automation allows a larger number of comparisons than any individual could perform.

The DD algorithm is used here to compare volatility surfaces, a description of the volatility across a financial market. Volatility signals the level of uncertainty in the returns of an investment and is commonly interpreted as risk or even fear by the investor. Often, the measure of volatility is represented by a number that is easy to compute and easy to compare in order to

understand and forecast the dynamics of volatility over trading days. However, such a simple representation may not be very informative because it reports only the average level of volatility and hides important differences in the location of volatility within the market.

The theory underlying volatility predicts that volatility will be constant across a market. After the market crash of 1987 traders observed a variation in volatility relative to the intrinsic value of a stock option and the length of time until its maturity (Derman, 1999). This lead to the proposal of more informative descriptions of volatility describing variation observed in different segments of the market. In addition to the Implied Volatility Surface (Cont and Da Fonsecca, 2001) measures such as volatility smiles, volatility skews, volatility smirks (Derman, 1999; Ayache *et al.*, 2004) have been proposed. Volatility smiles, volatility skews and volatility smirks provide theories of the mechanism for higher volatility at both ends of the maturity axis (Derman, 1999; Ayache *et al.*, 2004) while volatility surfaces offer a computational tool that traders can use to plot and compare the differences in volatility from a set of trades (Cont and Da Fonsecca, 2001). These representations provide more information and agree with the observations of traders, but make comparisons more complex. This chapter presents the use of deformation metrics as a simple but informative way to compare volatility surfaces.

One metaphor used to understand the comparison process is the idea of signal versus noise. Observations represent a noisy signal and the aim is to separate the noise from the underlying signal. The DD result contains a similarity or mapping distance and a mismatch term. Using the signal versus noise metaphor, the mapping between surfaces represents noise and the mismatch term is the signal of interest. For example, in MRI the calibration of machines and anatomical variation are noise, while anomalies in anatomical shape are the signal of interest. The standard procedures for analysing medical images involve initial adjustments to the image to account for

these noise sources. This is done by systematically deforming or morphing the image. Different deformations can be used depending on the type of noise present (Cachier *et al.*, 2003). Once the surface is morphed, any remaining distortion is of interest to the medical staff. The resulting deformed image is compared with the template image by a measure of distance between shapes. Both the deformations and the distance are measured in terms of energy. The overall metric is one that minimises the total deformation and distance energy (Beg *et al.*, 2005).

In this application to implied volatility surfaces, the signal of interest can be considered to be the shape of the volatility surface. The noise arises because there is incomplete information about the entire market on every trading day. Only parts of the market where trades took place are observed. Diffeomorphic mapping is used to compare the implied volatility surfaces and to develop a new index based on full information from the surface shape with the aim of detecting anomalous trading days where the shape, but not necessarily the level of volatility, is unusual.

9.1.1 Introduction to implied volatility surfaces

Volatility is variance of a price over time (Capinski and Zastawniak, 2003). True volatility can only be accurately measured retrospectively, but many other measures of the variation in price or anticipated variation in price are spoken of as volatility. In order to estimate future volatility a pricing model is used to assess current trades. The result is the implied volatility that the trader has built into the purchase price of options that will be redeemed for a set price on a future date. Implied Volatility Surfaces are a model which describes both the varying volatility of different trades and the varying volatility between trading days (Derman, 1999; Cont and Da Fonsecca, 2001; 2002; Cont *et al.*, 2002; Whaley, 2009). It is one of an explosion of volatility models that resulted from the Black–Scholes equation, which allows the accurate price of an option to be calculated, given volatility (Black and Scholes, 1973; Merton, 1973; Black, 1989; Hull, 2011).

Merton and Scholes received the Nobel Prize in Economics for their work (The Royal Swedish Academy of Sciences, 1997). The Black–Scholes equation changed options trading from an instinctual art into a quantitative exercise (Harford, 2012). Theoretical models have described the properties of volatility smiles, volatility skews and volatility smirks (Derman, 1999; Ayache *et al.*, 2004). Market analysis describes the changing volatility regimes over time (Hardy, 2001; Calvet and Fisher, 2004). An assortment of alternative models for the pricing of options have been proposed, such as jump-diffusion, latent volatility and nonlinear diffusion (Ankudinova and Ehrhardt, 2008; Aguilar, 2009; Psychoyios *et al.*, 2010; Doz and Renault, 2006), but the market continues to rely on Black–Scholes volatility. Volatility Surfaces can help bridge the gap between theory and practical application (Cont *et al.*, 2002).

Volatility signals the level of uncertainty in the returns of an investment. If an investment is free from any risk or uncertainty, and has a set price (strike price) at a set maturity date, then the fair asset price to pay today for the investment can be calculated with a simple compound interest formula such as

$$K = S\,(1 + r)^{T-t}, \qquad [9.1]$$

where $T - t$ is the length time to maturity, S is the asset price, K is the strike price and r is the risk-free interest rate. The compounding interest formula accounts for the time value of money. Risk-free investments are equivalent to bank deposits or investments in government-backed securities. If the asset price is put into an interest-free investment until the maturity date, the full strike price will be paid on the maturity date. If, on the other hand, there is a risk that the investment will not be paid on the maturity date, then the purchaser will reduce the asset price to account for this risk. The higher the risk of non-payment, the higher the risk premium deducted from the asset price (Capinski and Zastawniak, 2003).

To understand the risk of a market instead of that of an individual trade a proxy of the entire market is required. A proxy for an entire market is the trades made on an index fund comprised of representative stocks from that market. For example, the Standard and Poors 500 (S&P500) is comprised of 500 companies with large market capitalization selected to be a total market proxy for the New York Stock Exchange and the NASDAQ Stock Market (Standard & Poor's Financial Services LLC, 2012). Direct investment in the index is not permitted, but the Chicago Board of Exchange offers trades on several options based on the published value of the S&P500 index. For example, the SPX Option (ticker symbol SPX) is settled on the 3rd Friday of each month based on the published value of the S&P500 (Chicago Board Options Exchange, 2012).

The SPX Option is traded as call and put options. The purchase of a European call option allows the buyer to purchase an investment at a future maturity date and at an agreed strike price. The purchaser is obliged to sell, but the buyer is not obliged to make the purchase. Consequently, if the price of the underlying investment is less than the agreed strike price, the option will not be exercised. The opposite of a call option is a put option. A put option requires the buyer to purchase at the agreed price, but does not oblige the purchaser to sell (Capinski and Zastawniak, 2003). The price of a call option at maturity is

$$P(T) = max(S(t) - K, 0), \qquad [9.2]$$

and a put option is

$$C(T) = max(K - S(t), 0). \qquad [9.3]$$

where T is the maturity date, $P\ (t)$ is the price of the put option, $S\ (t)$ is the asset price, K is the strike price and $C(t)$ is the price of the call option. The price of the options at a date prior to the

maturity date is determined by the fundamental theorem of asset pricing (Delbaen and Schachermayer, 1994). An arbitrage situation exists when a trader can make profit without any risk (Capinski and Zastawniak, 2003). The fundamental theorem of asset pricing states that no arbitrage situations exist, and therefore the price is dependent on the value of the underlying asset and the time value of money. Using the trade information about all outstanding contracts, an investment can be derived that is assured to cover all the profits and losses of those outstanding contracts. Using these contracts and the final price of the asset, the price of the options is determined. Given a set of scenarios, Ω, the probability of each scenario, $\omega \in \Omega$, has a positive probability, $P(\omega) > 0$. The asset price discounted for the time value of money is $\tilde{S}_j(t) = S\,(t)/A\,(t)$, where $A(t)$ is a risk-free investment. The discounted options satisfy, $E\left(\widetilde{S_j}(T)\middle|S(t)\right) = \tilde{S}_j(t)$ for $j = 1, \ldots m$, the conditional expected value.

In order to price options, the price at the maturity date is required. Trades are made without this knowledge; therefore each trade can be viewed as an indication of that trader's assessment of volatility and each trade can be reverse-engineered to find the volatility implied by this trade. The Black–Scholes equation is a Partial Differential Equation (PDE) which uses Brownian motion to model the price of an option over time (Black and Scholes, 1973). A PDE is an equation that contains multivariate unknown functions and their partial derivatives (Boyce and DiPrima, 1997). Brownian motion describes the seemingly random motion of particles caused by collisions with smaller unseen particles (Einstein, 1956). Brownian motion equations have been used to describe seemly random motion of many natural objects such as animal movement and the diffusion of liquid (Codling *et al.*, 2008). The Black–Scholes equation is

$$\frac{\delta V}{\delta t} + \frac{1}{2}\sigma^2 S^2 \frac{\delta^2 V}{\delta S} + rS\frac{\delta V}{\delta S} - rV = 0. \quad [9.4]$$

where S is the price of the asset, $V(S,t)$ is the price of the index fund, t is time, r is a risk free rate of return and σ is the volatility. The equation quantifies the hedging required to remove all risk of volatility from a trade. It is often inverted to express a trade in terms of the implied volatility represented by the observed price, term and market condition (Cont and da Fonsecca, 2002).

To invert the Black–Scholes equation and use it to calculate the implied volatility of each trade, the observable components of the trade record required. The options trade record gives the contract term, the trade date, the strike price and the asset price. Each option is classified as either a call or a put. A call option gives the holder the right, but not the obligation, to purchase a stock at the contract price; in contrast a put trade gives the holder the right, but not the obligation, to sell a stock at the contract price (Capinski and Zastawniak, 2003). Options close to maturity are disregarded because their volatility is reflective of the options market as opposed to the underlying market of interest (Cont and da Fonsecca, 2002). A suitable risk-free rate for the date of the trade is required to make the Black-Scholes calculation. Following common practice, US treasury bonds are used as a risk-free investment because they are government backed securities. Combining this information, the Black–Scholes PDE can be solved for the implied volatility for all trades on the given date. The Black–Scholes formula for the call option is

$$C\,(S,t) \;=\; SN\,(d_1) \;-\; K\,e^{-r(T-t)}\,N\,(d_2). \quad [9.5]$$

The formula for the put option is

$$P\,(S,t) = \; K\,e^{-r(T-t)}\,N\,(-d_2) - \; SN\,(-d_1); \qquad [9.6]$$

where d_1 and d_2 are

$$d_1 \; = \frac{ln\,(S/K) + (r + \sigma^2/2)\,(T-t)}{\sigma\sqrt{T-t}},$$

$$d_2 \; = \; d_1 \; - \sigma\sqrt{T-t},$$

and $N\,(\cdot)$ is the cumulative distribution function, $T - t$ is the time to maturity, S is the asset price, K is the strike price, r is the risk-free interest rate and σ is the volatility.

Given the implied volatility of each trade, a summary of the volatility in a trading day was proposed by Whaley (1993). The summary was so well received that the Chicago Board Options Exchange (CBOE) developed an index based on the summary called the CBOE Market Volatility Index (VIX). The CBOE now trades futures and options based on the VIX index (Whaley, 2009). Because the VIX is a single number, it is easily graphed and compared over time.

Researchers have noted that implied volatility is neither constant over time, nor is it constant for all segments of the market at a given time. More informative models have been proposed to provide insight into the variation in implied volatility for a given day and the underlying market dynamics this represents (Derman, 1999; Cont and Da Fonsecca, 2001; 2002; Cont *et al.*, 2002; Whaley, 2009). One proposed model which presents the variation of volatility across a market and which complements the market practice of quoting options in terms of implied volatility is that of the volatility surfaces (Cont and Da Fonsecca, 2001). Volatility surfaces offer additional insight into the way that volatility varies across the market, but comparison between volatility

surfaces is not straightforward and differentiating between regimes is often done *a priori* for comparing shape changes (Derman, 1999; Cont and Da Fonsecca, 2001). Volatility surfaces have been compared using Karhunen–Louvre decomposition{Cont, 2001 #106}, a generalization of Principal Component Analysis (PCA), which describes the structure of sets of volatility surfaces, but does not make a comparison among them. This thesis proposes a metric to select anomalous volatility surfaces for further analysis.

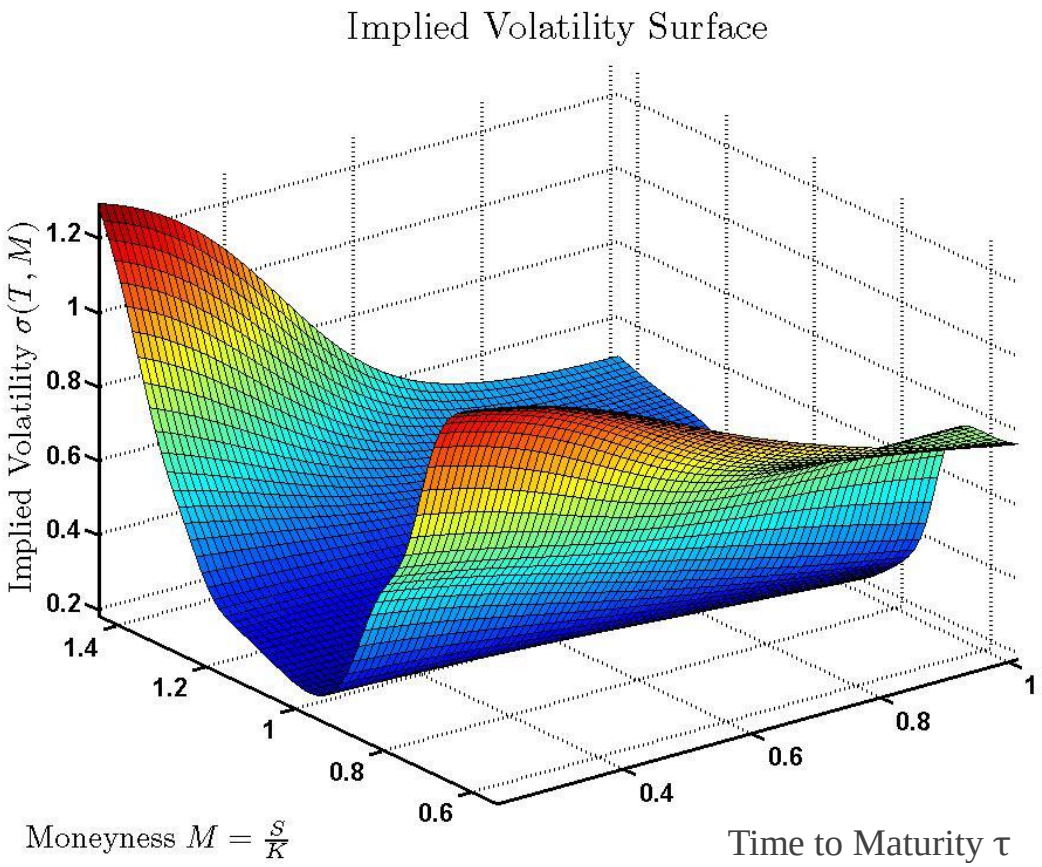

Figure 9.1: Example of an implied volatility surface for the S&P500. The axes are moneyness, time to maturity, and implied volatility.

An example of a volatility surface is shown in Figure 9.1. Volatility surfaces are expressed in a three-dimensional space, $\mathbb{R}^3$. The axes are time to maturity, τ, moneyness, M, and the implied

volatility, σ. Time to maturity is the number of days between the trade date and the maturity date, $\tau = T - t$. Moneyness is the strike price divided by the asset price on the contract date, $M = K/S_t$. Moneyness equal to one, $M = 1$, is called "at the money"; this means that the strike price is equal to the price of the underlying asset. "At the money" means that the option has no intrinsic value. Moneyness greater than and less than 1 are referred to as "out of the money" and "in the money" respectively (Capinski and Zastawniak, 2003). Surfaces are restricted to trades with moneyness values between 0.5 and 1.5 because trades are rare outside of this range and consequently a surface cannot be mapped accurately (Cont and Da Fonsecca, 2001). Implied volatility, σ, is calculated with the Black–Sholes PDE described earlier. From these observed trades, a kernel smoothing technique is used to estimate a smooth volatility surface in the three dimensions, M, τ and σ.

Each trade's implied volatility is a single observation about the shape of the implied volatility surface. In order to construct the entire surface, the product kernel estimator described by Scott (1992) and Silverman (1986) has been selected. Using the trades from a single day, at each point on the surface, (M, τ), the value of the implied volatility surface can be calculated using the formula

$$iv_d(M,\tau) = \frac{1}{n\, h_M\, h_\tau} \sum_{i=1}^{n} \left\{ K\left(\frac{x_M - x_{(M,\tau)}}{h_M} \right) K\left(\frac{x_\tau - x_{(M,\tau)}}{h_\tau} \right) \right\}, \qquad [9.7]$$

where d is the trading day, n is the number of trades in the trading day, x is an observer trade, and h is the bandwidth set using the Normal Reference Rule $h_j = \left({}^{4}/_{4n}\right)^{1/6} var(j)$. The resulting surfaces represent the implied volatility in the market, including combinations of M and τ that did not have any trades.

In addition to the surfaces for individual trading days, a reference surface is constructed. The reference surface is a composite of all the trading days and represents a "normal" day of trading. All trades are used to construct the reference surface. Given a suitably long trading history, the law of large numbers says that the resulting surfaces should be a better approximation of the expected value than each of the single trading day surfaces. The reference surface, iv_0, is constructed using the implied volatility surface formula and the normal reference rule, and all the trade in the time period,

$$iv_0(m,\tau) = \frac{1}{n\, h_m\, h_\tau} \sum_{i=1}^{n} \left\{ K\left(\frac{x_m - x_{(m,\tau)}}{h_m} \right) K\left(\frac{x_\tau - x_{(m,\tau)}}{h_\tau} \right) \right\}. \qquad [9.8]$$

where n is the number of trades over the entire period and h is the bandwidth set using the Normal Reference Rule stated above.

One surface is constructed for each day; this collection of surfaces is referred to as IV. The trading days are indexed using $d \in [1, D]$, where D is the number of days in the reference period. The trading day is referred to as IV_d . The reference surface, IV_0, represents the entire trading period in one surface.

The volatility surface for a given trading day describes the varying volatility across a market for the given day by using an index fund as a surrogate for all the stocks in the market. The volatility implied by each of the contracts traded for the index fund are used as observations of the implied volatility of a given moneyness and maturity. These are observations that are smoothed into a surface describing the volatility across all values of moneyness and maturity that comprise the market. Surfaces are constructed for each trading day and for the period as a whole. These smooth surfaces are discretised and the values on a regular grid are stored as a matrix of

computational purposes.

9.1.2 Diffeomorphic Demons literature review

The DD algorithm is a nonparametric diffeomorphic image registration proposed by Vercauteren *et al.* (2007; 2009). It is an extension of the Demon Algorithm proposed by Thirion (1998) and further examined by Pennec *et al.* (1999). Thirion's Demon Algorithm performs image-to-image matching and builds on his earlier work (Thirion, 1994; 1995). The foundation of both algorithms is Maxwell's demons, a concept describing diffusion across a semi-permeable membrane first studied in the 19th century.

After Thirion, diffusion methods became a fruitful area of research. Extensions have been offered: to improve convergence, stability and speed (Wang *et al.*, 2005); to develop a theoretical framework and to improve the force calculation (Cachier *et al.*, 1999); to allow for differing modalities (Kroon and Slump, 2009) and to adapt demons to the space of diffeomorphic transformations (Vercauteren *et al.*, 2009). The modality of data is the number of central tendencies it contains, observed as local maxima in the pfd. Diffusion methods show promise for offering good results relative to computation time (Hernandez *et al.*, 2008).

Maxwell's demons describe an apparent paradox in thermodynamics. Consider two gasses, A and B, in equal concentrations placed in two chambers, a and b, separated by a semipermeable membrane. The membrane contains 'demons' which allow gas A to diffuse from chamber b to chamber a only, and gas B to diffuse from chamber a to chamber b only. Over time, the concentration of gas A in chamber a increases and the concentration of gas B in chamber b increases (Figure 9.1). The entropy in each chamber appears to decrease. This contradicts the second law of thermodynamics, which states that the entropy of a closed system increases over

time. The resolution to the contradiction is that the demons require energy to distinguish between the particles (Leff and Rex, 1990; McGraw-Hill Concise Encyclopedia of Science and Technology, 2005).

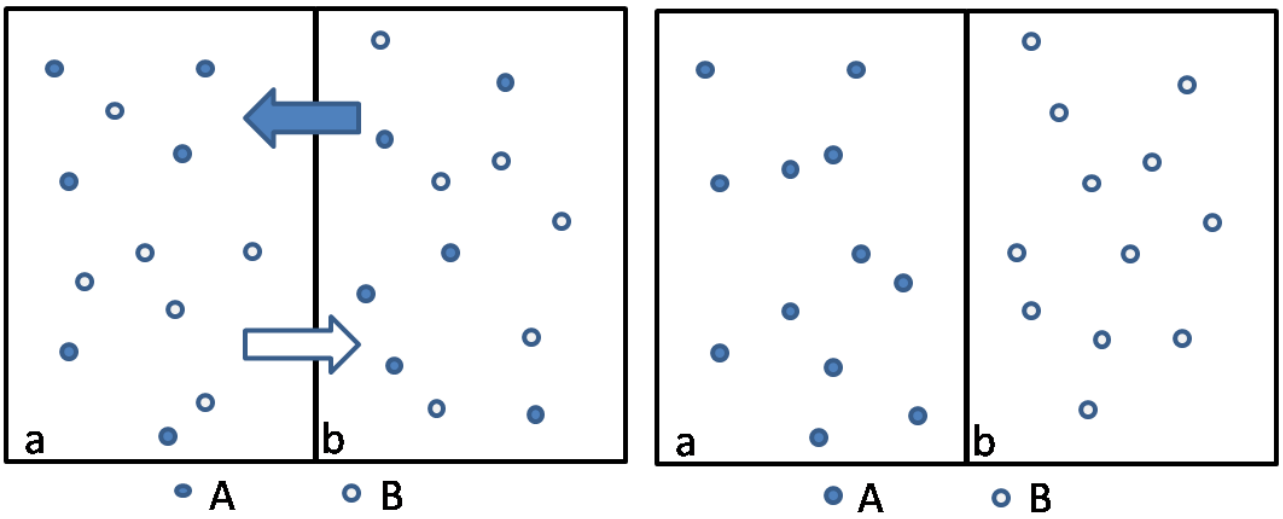

Figure 9.2: Maxwell's demons illustrative example. The figure on the left has equal concentration of gas A and gas B. The membrane allows each gas to diffuse in only one direction. Over time, the system will move towards the figure on the right.

In DD, the notation and naming of source and target image (I_0 and I_1) used in LDDMM are replaced by moving image, and fixed image (M and F), but the roles are the same. The moving image is transformed and the fixed image directs the transformation. The fixed image acts as the membrane and the demons are placed on this membrane to control its movement. The result of the movement is a moving image more like the fixed image. Each demon can act in one of two equal and opposite directions. Demons will either "push" or "pull" the moving image so that it aligns with the fixed image. The decision to "push" to the outside from the surface or "pull" towards the interior of surfaces is based on polarity. The polarity for each demon is determined by the location of the demon relative to the moving image. Each demon is either inside or outside of the moving image. Using the polarity, demons determine which direction to act

(Thirion, 1998).

The Demon Algorithm detailed by Thirion is a general framework of methods encompassing several versions suited to different situations. Demons operate on the set of all allowed transformations $\mathcal{T}$ between the space $\mathcal{M}$ of the moving image M and the space $\mathcal{F}$ of the fixed image F. The method iterates over a series of transformations, $T_0, T_1, T_2, \dots$, to arrive at a final transformation, $T \in \mathcal{T}$, comprised of all the iterations $\{T_0, T_1, T_2, \dots\} \subset T$. Each iteration is viewed as a time-step and at time i, $T_i(M)$ becomes $T_{i+1}(M)$ by the action of two forces, the internal force, f_{int}, and the external force, f_{ext}. The internal force, f_{int}, is dictated by the points on M, whereas the external force, f_{ext}, is dictated by the point interaction between the moving image M and the fixed image F at the beginning of the iteration $T_i(M)$.

The general framework of all demon algorithms is as follows:

- **Step 1.** Place all the demons on F and pre-compute their values, D_s.
- **Step 2.** At time $+1$, compute the forces f_{ext} and f_{int}, given T_i (M) and F.
- **Step 3.** Compute the transformation T_{i+1} from the forces f_{ext} and f_{int}, and the result of the prior iteration T_i (M).
- **Step 4.** If stopping criterion has been met, $T = T_{i+1}$; otherwise return to Step 2.

The first step is the initialisation step. There needs to be a placement scheme for the demons. The common methods are to place one demon for each pixel/voxel or to place demons only at contour points of F (Thirion, 1995). Each of these methods yields a set of points on the surface P on D_s. Next, for each demon, the direction $\vec{d}$ from the inside to the outside is determined. This is generally done with the gradient of the surface, $\vec{\nabla} s\,(P)$. Finally, the displacement from M to F is calculated using $\vec{d} = \overrightarrow{PP'}$, where $P' = {T_i}^{-1}(P)$. The distance is measured on some interface information such as the value of the intensity at $s\,(P)$.

Steps two and three are the iterative steps, and Step four is a test to see if the iterations can stop. The initial value of the transformation is the identity mapping $T_0(M) = M$. In Step two, at iteration i, the demon force, f_i, is computed at each demon's position $P \in D_s$. This is done based on the polarity of M at $T^{-1}(P)$ and the demon's direction $\vec{d}$ calculated during the initialization. In Step 3, the force for each demon in Step 2, $(f_i(P), P \in D_s$, is applied to T_i and the result is T_{i+1}. Step 4 uses some similarity measure to test for convergence of the transformed moving image and the target fixed image, $sim(\mathrm{T_{i+1}(M)}, \mathrm{F}) < \epsilon$.

In medical applications the Demon Algorithm has been shown to have several utilities. It has high registration precision when used to measure the subthalamic nucleus region of the brain (Castro *et al.*, 2006). When it is used for comparing brain scans of two different patients the anatomical features that are similar are aligned, while those that differ do not align and remain after the transformation (Thirion, 1998). There are some issues in this algorithm: images with differing modalities cannot be matched. An extension that allows for differing modalities is offered by (Kroon and Slump, 2009). Another issue is that images without any initial overlap between them cannot be matched.

Researchers have extended the demon framework and applied it to the space of diffeomorphisms (Vercauteren *et al.*, 2007; 2009). DD are guaranteed to reach convergence via a geometric optimisation. The original Demon Algorithm was on the space of all displacement fields and the resulting displacements were vectors. Successive displacement vectors are added to get the final transformation. When using the space of all diffeomorphisms, Lie groups are used and successive transformations are applied by composition to get the final transformation.

DD uses a nonparametric transformation to points $s(.)$ to map points on the fixed image space,

p, to points on the moving image space, $s(p)$. The transformation is described by a displacement field $\boldsymbol{s}$ which when added to the identity transformation results in the desired transformation. The stopping criterion or convergence test for DDs is the mean square error between the intensity of the two images

$$Sim\left(F, \dot{M} \circ s\right) = \frac{1}{2}\|F - M \circ s\|^2 = \frac{1}{2|\Omega_P|}\sum_{p\in\Omega_P}\left|F(p) - M\left(s(p)\right)\right|^2, \qquad [9.9]$$

where s is the transformation, $\Omega_P$is the region of overlap between F and $M \circ s$. A regularization term is added to the global energy of the transformation s to ensure stable and smooth solutions.

$$E(s) = \frac{1}{\sigma_i^2} Sim(F, M \circ s) + \frac{1}{\sigma_T^2} Align(s), \qquad [9.10]$$

where s is the transformation from the fixed image to the moving, $Sim()$ is the function in [9.9], $Align(s)$ is a regularisation term, and $\frac{1}{\sigma_i^2}$ and $\frac{1}{\sigma_T^2}$ are weighting terms. The formula for regularisation is $Align(s) = \|\nabla s\|^2$. One further source of uncertainty is introduced to the global energy during computation, to simplify the optimization step. The transformation c is introduced to allow for some error between the result of the spatial transformation s and the pixel values of the moving image. The nonparametric transformation c is a vector field containing point correspondences between image pixels. This results is a global energy which is optimized by the DD algorithm

$$E(c, s) = \frac{1}{\sigma_i^2} Sim(F, M \circ c) + \frac{1}{\sigma_x^2} dist(s, c)^2 + \frac{1}{\sigma_T^2} Align(s). \qquad [9.11]$$

Where the first term enforces matching between the two images, the second term allows for

uncertainty in the match by adding Gaussian noise to the match, and the third term is a hidden regularization term. The third term is introduced to add *a priori* knowledge about the match and ensures that this is a well posed problem with smooth results. The σ_i^2 term controls the measurement error within each image, σ_x^2 controls the matching error between images, and σ_T^2 controls the level of a hidden factor called the regularisation of the images. The formula used for distance is $dist\ (s,c)\ =\ \|c-s\|$. The formula for regularisation is $Reg(s) =\ \|\nabla s\|^2$.

This leads to the approximation of $E(c,s)$

$$F(p) - M \circ s \circ exp(u)\,(p) \approx F(p) - M \circ s(p) + J^p.u\,(p), \qquad [9.12]$$

where u(p) is the update field

$$u(p) = -\frac{F(p)-M\circ s\,(p)}{\|J^P\|^2+\frac{\sigma_i^2\,(p)}{\sigma_x^2}} J^{P^T}\ , \qquad [9.13]$$

$J^p = -\nabla_p^T\ (M \circ s)$ or $J^p = -\nabla_p^T F$, and $d(s,c) = \|Id - s^{-1} \circ c\|$. Further leading to the distance formula $dist(s, s \circ exp(u)) =\ \|Id - exp\ (u)\| = \|u\|$. The diffeomorphic energy can be restated in the demon energy formula

$$E_s^{corr}(u) \approx \frac{1}{2|\Omega_P|}\Sigma_{p\in\Omega_P}\left\|\begin{bmatrix}F(p)-M\circ s(p)\\0\end{bmatrix}+\begin{bmatrix}J^p\\\frac{\sigma_i(p)}{\sigma_x}I\end{bmatrix}.u(p)\right\|^2. \qquad [9.14]$$

Following Vercauteren (2007), the steps for DD computation are:

Step 1. Compute the correspondence update field u using [9.13].

Step 2. Let $u \leftarrow K_{fluid} \star u$, if using fluid-like regularisation.

Step 3. Let c ← s ∘ exp (u), with the equation for fast computation of the vector field exponent algorithm.

Step 4. If a diffusion-like regularisation is used, let $s \leftarrow K_{diff} \star c$; otherwise let s ← c .

The fast computation of the vector field exponent algorithm is:

Step 1. Choose N such that $2^{-N}u$ is close enough to 0, e.g. $max\|2^{-N}u\ (p)\| \leq 0.5$.

Step 2. Perform an explicit first order integration: $v\ (p) \leftarrow 2^{-N}u\ (p)$ for all pixels.

Step 3. Do N recursive squarings of v: v ← v ∘ v.

The results of the DDs yield smoother spatial transforms than demons, but with a penalty of a 50% increase in computation time per iteration (Vercauteren *et al.*, 2007).

The DD algorithm offers a computationally efficient image matching algorithm based on diffusion methods. The diffusion is controlled by demons placed on one surface. The demons guide the diffusion so that the scene image is deformed to look like the model image. The demon method contrasts with the other image matching techniques that use an attraction or similarity methods to determine the movement. Attraction methods are computationally expensive, as high as $O\ (n^2)$. The LDDMM method discussed in the previous chapter is an example of an attraction method. It uses a force based on similarity to determine the path of the movement (Thirion, 1998). LDDMM requires a computationally expensive PDE solution during each iteration (Vercauteren *et al.*, 2007). One drawback of the DD algorithm not found in diffusion methods is that all surfaces are required to have the same modality. An extension has been proposed to handle changes in modality (Kroon and Slump, 2009).

9.2 Method

9.2.1 Diffeomorphic Demon calculation

The first step in applying DDs to the implied volatility surfaces is to map each individual day's surface to the reference surface (the comparison surface) by a shift on the moneyness and maturity axes. The deformation mapping is done while preserving the volatility values and the relative point positions, thus preserving neighbourhood properties. Demons Algorithm (Thirion, 1998) takes inspiration from models of particle diffusion studied in thermodynamics. The effect of this when applied to two images is that one image controls the diffusion of the other (Thirion, 1998). The application aligns like neighbourhoods in the volatility surfaces. In medical images, motion is restricted to real space and comparisons are made based on image intensity, with implied volatility motion being restricted to the moneyness and time to maturity axes, and comparisons are made on the implied volatility axis.

The aim is to find individual trading days that have a shape that is an anomaly when compared with other trading days. In order to do this, the shape of a typical trading day needs to be decided. In computational anatomy, practitioners have developed a "reference brain" that identifies all the typical shapes and locations in a human brain. That way, other brains can be mapped to this reference brain. There are several theoretically typical shapes for the volatility in the market that are candidates to act as a theoretical reference, e.g. the volatility smile and the volatility smirk (Derman, 1999; Ayache *et al.*, 2004). These models describe an increase in volatility for in and out of money options, but do not offer a formula to derive the expected volatility given a set of market conditions. For this reason a theoretical volatility shape has not been used here; instead a reference volatility surface has been constructed by using all trades

over the reference period. This will give the composite shape of the volatility surface over the entire period. As long as the reference period is suitably long, the anomalous days should make only a very small contribution to the shape of the composite surface. By definition, anomalous days should be few and balanced by normal trading days. If the reference period is very short, an alternative method to prevent the trading from the day of interest from influencing the shape is to use a hold-one-out method to calculate a different surface for each trading day using all the trades except those that occurred in the given day.

A composite or template implied volatility surface is constructed by using all the trades in the entire period. This template acts as the target for deformations by aligning the implied volatility surface of interest to this target. The surface is deformed with the Demon Algorithm. Here a point-wise comparison is made to the template. This comparison is a similarity metric that is based on the error term of the deformation. The combined energy required for deformation and comparison is referred to as the deformation metric. It represents the shape difference between the surface of interest on a given day and the composite template. It is a direct measure of shape difference. Higher deformation metrics are associated with the more anomalous implied volatility days.

9.2.2 Mechanism of DDs: surface comparison

This section describes how the deformation metric is constructed and demonstrates its utility by applying it to information from a series of trading days to identify anomalous days. In order to compare the shape of the surfaces, a similarity metric, $Sim\ (IV_d, IV_0)$ is constructed to quantify the resemblance of the two surfaces. The metric compares each point on a trading day's surface with the corresponding point of the reference surface. The corresponding point is the end point of the mapping curve that connects the trading day surface to the reference surface: E_{sim} (IV_0,

$IV_d) = |IV_0 - IV_d|^2$.

A high value of the similarity index E_{sim} indicates high dissimilarity and a low value indicates similarity. Identical surfaces have $E_{sim} = 0$. Because of the squaring of the results, large localised differences between the models will have a higher index than small global differences between the models. Thus the shape difference is captured, not simply distance.

This similarity metric cannot be applied directly to IV_0 and IV_d, because if the surfaces are not perfectly aligned then the dissimilarity will be exaggerated. For example, if two identical surfaces were compared, the similarity would be zero. Now if one of these surfaces was shifted on the moneyness axis, the shape would remain the same, but this would create distance between corresponding points, thus increasing the similarity index. To avoid this, the surfaces are aligned with a deformation before comparison with the Demon Algorithm (Ayache, 2007).

The alignment process requires IV_0^* ; this is IV_0 aligned to IV_d. The alignment is performed using a deformation of IV_0 that allows IV_0 to move in the M and τ axes while fixing the σ values. The surface IV_d acts as a target to align the landmarks, peaks and valleys, of IV_0. The magnitude of the transformation is measured with the energy needed to perform the alignment

$$IV_0^* = IV_0 \circ \varphi$$

$$E_{align}(\varphi) = |\nabla \varphi|^2_{fro}., \qquad [9.15]$$

The total difference between the surfaces is the alignment energy, also called the similarity energy. There are many possible alignments and similarity combinations for any two surfaces, the aim is to find the combination which minimise the joint energy. Weightings $1/\sigma^2_{align}$ and

$1/\sigma_{sim}^2$ are introduced to keep the alignment and similarity energies in balance

$$E(\varphi) = \frac{1}{\sigma_{align}^2} E_{align} + \frac{1}{\sigma_{sim}^2} E_{sim}, \quad [9.16]$$

where

$$E_{sim}(IV_0^*, IV_1) = |IV_0^* - IV_1|^2,$$

E_{align} is the alignment energy, and φ is the diffeomorphic transformation. If the weight of the alignment energy is too large, then IV_d^* will not resemble IV_d. If the weight of the similarity measure is too high, then IV_d^* and IV_d will be nearly identical and the dissimilarity will be exaggerated.

To estimate the transformation, φ, The source image iteratively moves the source model towards the target model. At each step, the direction of the transformation is estimated by calculating the extended demon force (Joshi *et al.*).

To estimate the transformation, φ, the scene image is iteratively moved towards the model image. At each step, the direction of the transformation is estimated by calculating the extended demon force (Joshi *et al.*, 1995)

$$force = (IV_0 \circ \varphi - IV_1)\left(\frac{\nabla IV_1}{|\nabla IV_1|^2 + \alpha^2 (IV_0 \circ \varphi - IV_1)^2} + \frac{\nabla IV_0}{|\nabla IV_0|^2 + \alpha^2 (IV_0 \circ \varphi - IV_1)^2}\right), \quad [9.17]$$

where α is a noise constant and φ is the transformation calculated on the prior iteration. The

force is then smoothed by a Gaussian kernel and added to the transformation calculated on the prior iteration. The demon force pushes each point on the moving surface in the direction determined by the gradients of both surfaces. The size of each step is determined by the remaining difference between the moving image at the last time step and the static model image. The resulting transformation is a vector field, φ, that maps each point on IV_0 to a new location on the aligned surface, IV_0^*.

9.3 Case Study

A comparison of Implied Volatility Surfaces is performed using the S&P500 index options traded on the Chicago Board of Exchange (ticker symbol SP500) from 1 January 2008 to 27 February 2009. The SP500 options are traded with European-style call and put options. One surface was constructed for each day of trading with axes for moneyness and time to maturity for a total of 291 surfaces. A reference surface was constructed using all the trades in the period. Each daily surface was then compared to the reference surface.

To remove additional volatility in options close to maturity, options within two months of maturity are discarded. Trades with $M > 1$ were treated as calls and $m \leq 1$ were treated as puts. The US Federal Reserve 1-year T-bill rate was used for the risk-free interest rate.

An example of a single day's surface before and after this transformation is shown in Figure 9.4. This example shows that the transformation preserves neighbourhood relationships by the fact that the mesh of the transformation does not cross itself. Both the original and the aligned Day 175 surfaces have peaks in the lower left-hand and lower right-hand corners, but in the aligned surface, both peaks have spread upwards and to the centre.

Looking at the time series in Figure 9.3 two days are drastically different to the remainder of the period: Day 99 (22 May 2008) and Day 50 (13 March 2008). Although the level of volatility increased in the second half of the time series, the shape of the volatility surface became less variable in this period.

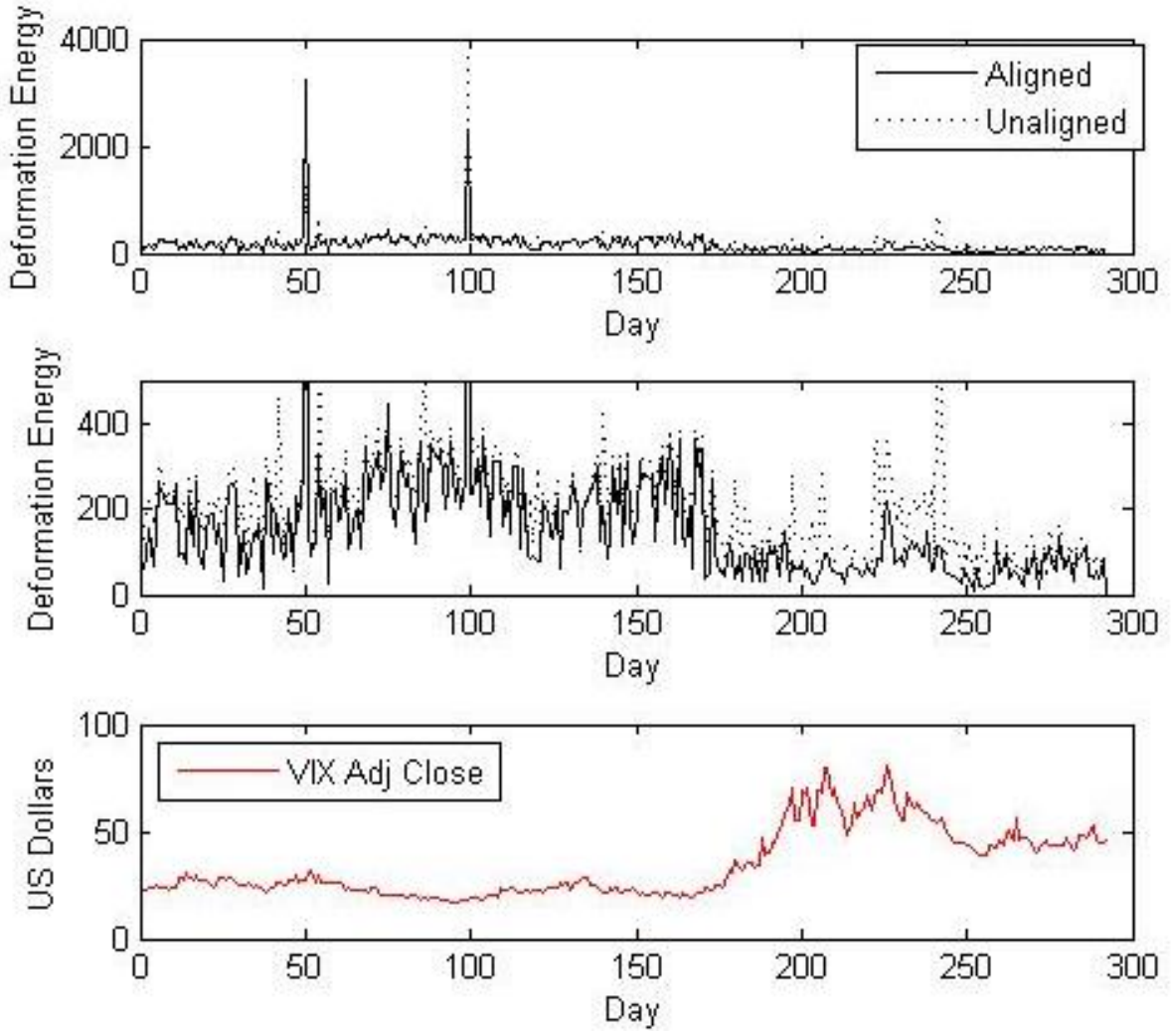

Figure 9.3: Metric mapping with alignment and similarity contrasted with the similarity without alignment, and the volatility index (VIX) over the same period of time.

Figure 9.5 shows three volatility surfaces: the reference surface, an anomalous surface and a typical surface. Comparing the anomalous Day 99 (Figure 9.5 (c) and Figure 9.4) to reference surface (Figure 9.5 (a)) and the more typical Day 175 (Figure 9.5 (b)), the shape of the Day 99

surface is very different from the other two surfaces. This indicates that there was volatility in a different part of the market on Day 99 than on a normal trading day. The features on day 99 – flat volatility everywhere except low moneyness and high time to maturity – are not seen in the other surfaces. The shape differences are not due to an anomalous number of trades. The average number of trades per day is 375 over the period and all days have a sufficient number of trades to accurately define the IV surface. Intuitive shape comparison says that Day 175 is closer in shape to the reference than Day 99. This could be of potential use to detect early market signals. Future work will investigate this possibility.

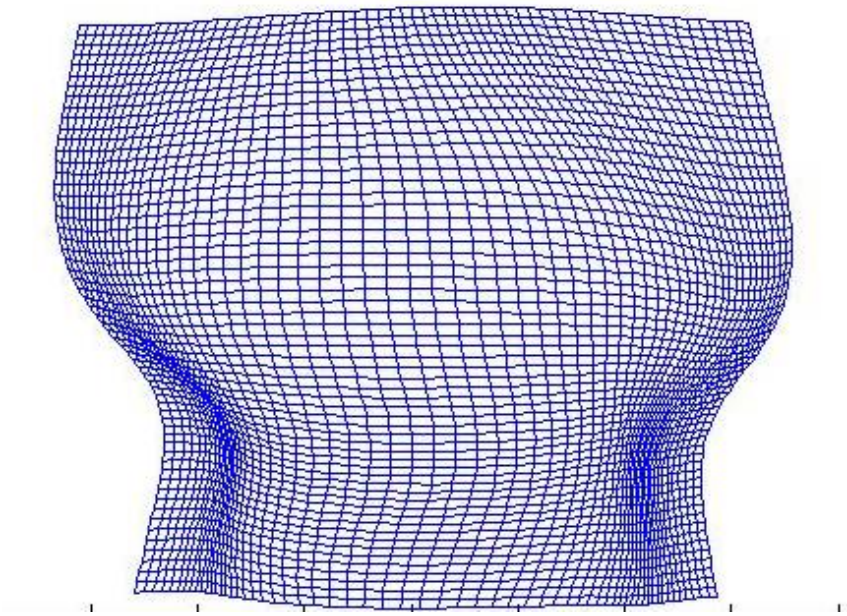

Figure 9.4: The transformation of Day 175. The surface for Day 175 was aligned to the surface for the entire period.

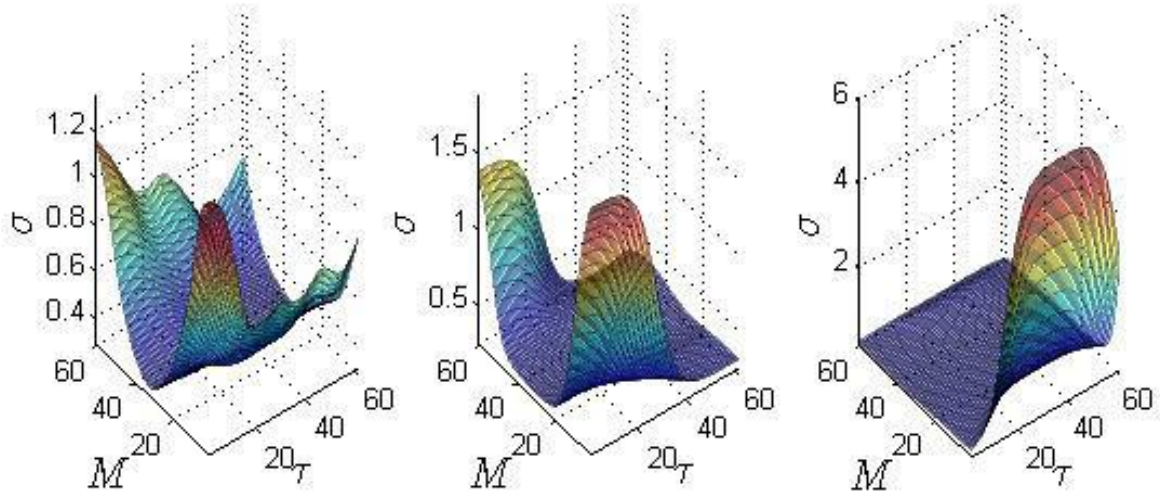

Figure 9.5: (a) Implied volatility of SP500 over the entire period (from 1 January 1 2008 to 27 February 2009. (b) The implied volatility surface for Day 175. (c) The implied volatility surface for Day 99.

The intuitive sense of closeness is collapsed into a single number, thus making comparison simple and allowing a large number of surfaces to be compared in a consistent manner. In addition, the constituent parts of the metric (the aligned figure and the vector field of the deformation) can offer us additional insights into the locations of the shape differences.

9.4 Discussion

Since the volatility skew appeared in the market, implied volatility has been analysed and predicted, beginning with apocryphal rules that did not hold (Derman, 1999) and moving to accurate models. However, there are still fundamental open questions about the nature and cause of the skew (Ayache *et al.*, 2004). This method of comparison is another tool to describe market volatility over time. It allows comparisons based on the shape of the volatility surface. This additional shape information supplements the information provided by the VIX that is widely used to track the level of volatility with a measure of unusualness of the location of the volatility as well as its level.

A possible explanation for the two anomalous days, 5 May 2008 and 13 March 2008, is that both days saw record oil prices. In addition, on May 5th, the US Federal Reserve released a report to banks about tightening lending practices, and on the 13th of March, the US Dollar reached a 12-year low against the yen.

This type of shape comparison has thus far been restricted to comparing images in a medical setting. Volatility surfaces are less tangible than medical images, but professional traders are interested in shape changes that indicate market conditions changing, just as medical professional are interested in shape changes that indicate disease.

9.5 Conclusion

In this chapter the Diffeomorphic Demons algorithm originally formulated for computational anatomy was applied to a new context. The new context is comparing implied volatility surfaces for a series of trading days in a stock market. A similarity metric to quantify the shape difference of the volatility surface for trading days was used to detect anomalous trading days. A large set of trading days was compared using an automated method to give the same results as intuitive comparison for shape difference.

Volatility surfaces have been proposed as a more informative measures of market volatility than weighted averages such as the VIX that would help traders make more informed decisions (Cont and Da Fonsecca, 2001) a likely complexity that has hindered wider adoption of volatility surfaces is the difficulty in making surface and shape comparisons. Volatility surfaces offer benefits to market practitioners because they describe the variation in volatility dependent on moneyness and maturity observed by traders. The volatility of a market is converted into a volatility surface by using an index fund based as a proxy for the stock trades in the market. The

volatility of each option contract made on the index fund is converted into its implied volatility using the Black-Scholes equation. These observations of implied volatility are then combined with kernel smoothing and the result is a surface that describes the implied volatility for all values of moneyness and maturity.

The comparison of volatility surfaces with DD can add to a trader's understanding of market forces. In stock market trading the more understanding a trader has of market dynamics, the greater advantage they have over other traders and the more money they make. As an example, when Black-Scholes was first employed the traders who used it worked alongside intuitive trading. The quantifiable understanding of market forces provided by Black-Scholes put the professional traders not using it at a disadvantage and within a short time Black-Scholes became universally adopted (Harford, 2012).

Diffeomorphic Demons is an iterative method that measures the distance between surface by measuring the energy needed to deform one surface into another. DD avoids the computationally expensive task of solving a PDE at each iteration of the algorithm required by other deformation metrics. The PDE is avoided by using diffusion equations. The diffusion is guided by demons placed of one of the surfaces. Demons either "push" or "pull" the surface so that it aligns with the model. The decision to "push" to the outside from the surface or "pull" towards the interior of surfaces is based on polarity. Each demon is given a polarity based of it being either inside or outside of the source image. This polarity determines which direction the demon will act.

9.5.1 Future Work

Future work in this area includes comparison of the performance of DD metric against the generalised autoregressive conditionally heteroskedastic models with a lag length of 1 and

autoregression order 1 (GARCH 1-1) commonly used to forecast volatility in markets (Hansen and Lunde, 2005). The GARCH model (Bollerslev, 1986) is a generalisation of the the autoregressive conditionally heteroscedastik (ARCH) models proposed Engle (Engle, 1982).autoregressive conditionally heteroskedastic (ARCH) models proposed Engle (1982). The ARCH and GARCH models describe market dynamics as a stationary process with an autoregressive structure and inconsistent variance over time.

Another direction for future work is replacing the template surface constructed from all the trades in the period with theoretical models for market volatility such as the volatility smile and the volatility smirk. The DD metric could then be used to determine which of the models more closely matches a given trading day. This comparison may detect switches in the volatility regime determined through expert knowledge by Derman (1999).This comparison may automate the detection of switches in the volatility regimes currently determined through expert knowledge (Derman, 1999).

10 Summary

Every observation in a statistical dataset is made at a time and a place. Spatio-temporal statistics explicitly includes these spatial and temporal components in the analysis of the data. This thesis presents selected topics is spatio-temporal statistics: removing bias from spatio-temporal models and comparing regression surfaces resulting from spatio-temporal models. In this summary an overview of the thesis is presented, a summary of each chapter given and some concluding comments are made.

Removing bias from habitat observations made with radio telemetry equipment is examined and two methods are presented. The parametric method models a biological process, animal movement, to impute missing data. The imputed data are used to adjust the dataset and these adjusted data are used for analysis. The non-parametric method for correcting habitat uses bias in radio telemetry studies uses a Bayesian paradigm. The analysis is done with a Bayesian Hierarchical Model (BHM) which performs the analysis and accounts for missing data without creating a bias adjusted dataset. In the Bayesian framework back and forth estimates are made starting with the prior estimates, a model for bias weights is estimated given observed data, then the likely observed data are estimated given the model for bias weights, again the model for bias weights is estimated given the revised data , etc. This iterative process is repeated until some resting state is reached.

The second section of the thesis presents comparisons of spatio-temporal data with deformation

metrics from computational anatomy. Two methods are presented: comparing statistical regression surfaces using Large Diffeomorphic Metric Mapping (LDDMM) and comparing the volatility of financial markets with Diffeomorphic Demons (DD). These are the first applications of LDDMM and DD outside of computational anatomy.

Computational anatomy had made great strides in comparing complex spatio-temporal observations, but the literature is not widely cited outside of this community and is mathematically dense. The complexity of the calculations and the rigorous mathematical presentation do not foster adoption in other areas of spatio-temporal statistics. Several novel analogies are presented for understanding the computational anatomy concepts to aid those unfamiliar with the area in understanding, applying the mathematical formulations, and analysing the results.

10.1 Chapter 6

Chapter 6 interpolates missing radio telemetry data for a single animal using an animal movement model. The missing values are modelled using characteristics of the animal's movement during non-missing periods. Two methods are presented: parametric and non-parametric. The parametric method applies animal movement models widely used in other areas to bias correction for the first time. The non-parametric proposes an animal movement model based on kernel smoothing.

In order to speed the computations when applying the parametric method the standard turning angle was replaced with a bearing angle. Both measure the change in direction along an animal's path, but the bearing angle offers computational gains. The computational gains are made because the bearing angle is fixed across the entire study area and the movement direction of the

animal does not need to be calculated at each time point. In addition, only a single successful fix is needed to apply the bearing angle calculation, whereas two fixes are needed to calculate a movement direction and apply a model based on turning angle.

The non-parametric method is computationally efficient because all of the distances and turning angle computations required to apply the parametric movement model are not required by the non-parametric movement model. The relative distances between successful fixes are calculated once while estimating the nonparametric movement model. The non-parametric computation in this application is simply matrix addition. The home range model is also converted to a matrix to speed application. Restraining the model to the home range is a simple calculation of the entrywise product of two matrices.

Modern radio tracking studies offer insight into animal behaviour at small time intervals and large study areas. The bias of the data collected by these studies is recognized, but the methods for adjusting to remove bias are still not widely agreed or applied (Frair *et al.*, 2004). The parametric and nonparametric models for estimating detection probability offered here estimate the bias for a single animal directly from the collected data instead of conducting separate studies to determine detection rates of equipment at each habitat type in the study area. The model builds on well researched methods of estimating animal movement and animal home range.

Two case studies are presented: a simulated dataset and a wild boar dataset. The parametric method is used on a simulated dataset to create a bias adjusted dataset. A resource selection function is then fitted with both the bias adjusted dataset and the unadjusted dataset and the results are compared. The nonparametric method is demonstrated with data from a wild boar study in southern Sweden. The study placed GPS collars on sows in the summer of 2005 and

recorded their location at half hour intervals (Thurfjell *et al.*, 2009). The method is demonstrated with a single boar's data (boar number 1473). This boar has 19.9% of her fixes missing over the tracking period.

Both case studies show that the method successfully adjusts for bias in the raw data of one tracked animal. The results of applying the method to a simulated dataset show that the bias introduced into the simulation is removed. The results of analysis performed on the bias adjusted boar datasets are consistent correcting bias in the raw data. The drawback of this method is scaling it to larger datasets containing multiple animals.

10.2 Chapter 7

Chapter 7 presented a new method for constructing regression surfaces for Automated Telemetry System (ATS) detection probability using a Bayesian Hierarchical Model(BHM). This is the first time that a Bayesian formulation of detection probability has been proposed for the correction of ATS bias. It reframes Johnson *et al.* (1998) method for estimating ATS detection probability using Bayesian methods. The BHM extends occupancy estimation models (MacKenzie *et al.*, 2002; MacKenzie *et al.*, 2005; MacKenzie, 2006; MacKenzie *et al.*, 2006; MacKenzie *et al.*, 2009). Occupancy models adjust for animal non-detection with the assumption that if the animal was present during any visit to a site the animal occupies the site for the entire study period. Because ATS tracks a single animal over multiple locations, information about the animal's presence at one location can make detection at another location impossible. The BHM for occupancy is adjusted by adding an additional parameter for presence required to make the model suitable for this context.

Johnson *et al.* (1998) considered one source of variation due to radio transmission and carefully

defined stationary periods to eliminate all other sources of variation. A stationary period is defined as a six hour window where the animal moves less than 200m. In this work I replace stationary periods with temporal and spatial neighbourhoods. The temporal neighbourhood of a fix is the three hours before and after the fix occurred and the spatial neighbourhood is the 30 by 30 grid square that contains the fix. This adjustment to Johnson *et al*. (1998) model allows more of the fix data to be used in the model fitting process.

The method is demonstrated with the same case study also used by Johnson *et al.* (1998), the Starkey Experimental Forest. In this dataset locations for 100 plus animals were collected for multiple years. Datasets of this size are rare and demand computationally efficient methods. The results of the new method are compared to those Johnson *et al.* (1998). The calculations were done using WinBugs (Lunn *et al.*, 2000) on a laptop, the estimates of the detection probabilities were completed in a reasonable amount of time (less than 4 hours) without expensive computing equipment.

10.3 Chapter 8

In Chapter 8 the LDDMM is used as a measure of global similarity between two regression surfaces. I developed a four step process to compare statistical regression surfaces with LDDMM. The steps are: the surfaces are aligned via landmarking, the aligned surfaces are discretised to give a computational grid conducive to accurate comparison, the surfaces are converted to images to simplify computation with image libraries in MATLAB (MATLAB, 2012), and finally a mapping is made with LDDMM.

One advantage over standard comparison is that the alignment allows like area of the surfaces to be compared. Standard comparisons measure change with the spatial location, but spatio-

temporal processes may make spatial alignment erroneous because of localised changes in the spatial environment, for example changing land use or localized drought conditions. Comparing LDDMM to other computational anatomy comparisons, it has many desirable properties. It measures distance with fluid flow equations. The distance is measured at a series of deformations of one of the surfaces. Each interim deformation is a timestep and the fluid flow equations used preserve features of the surface over many time steps. LDDMM also preserves neighbourhood relationships; the points close together on a surface are still close after the LDDMM mapping. Similarly points far apart on a surface remain far apart after the LDDMM mapping. The construction of the LDDMM mapping ensures that a backwards mapping always exists.

The LDDMM metric is explained using a simple two dimensional dataset, then demonstrated on two case studies of higher dimensions. The explanatory dataset is the wingspan of aircraft in three time periods of aircraft design. One case study is detection probability maps calculated in Chapter 7 for a radio telemetry study at Starkey Experimental forest in Oregon, USA. The other case study is measures of air quality collected in New York City. These case studies demonstrate that LDDMM shows promise in comparing spatio-temporal data with misalignment due to temporal changes in the study area.

10.4 Chapter 9

Chapter 9 uses DD from computational anatomy to compare a collection of regression surfaces and find those with anomalous shapes. This is the first application of DD outside of computational anatomy. The application is to compare financial data, specifically implied volatility surfaces. Implied volatility surfaces are constructed to describe the market volatility in

different segments of a financial market. One surface is constructed for each trading day and the surface describes the volatility as a function of moneyness (the intrinsic value of the financial instrument being traded) and the length of time until the maturity date.

The temporal nature of implied volatility surfaces is straightforward; there is one surface for each trading day. The spatial nature of implied volatility surfaces is not as straightforward. The two dimensions of moneyness and time to maturity are treated as the "location" in the market. Just as in physical locations there are neighbourhoods in the market and the distance between two trades can be measured on the moneyness and time to maturity axes.

The aim of this analysis is to capture and quantify changes in market conditions intuitively understood and widely discussed by market analysts and traders. The DD metric differentiates anomalous trading days where the shape of the volatility surface is different from typical trading days, even if the level of volatility is not anomalous.

A volatility surface is constructed for each trading day and a template Volatility surface is constructed by using all trades over the period being analysed. Then DD iteratively fits the deformation to each of the trading day surfaces and measures the energy required to deform the given trading day into the typical trading day.

10.5 Concluding comments

Many current statistics articles and books begin with an explanation of the fundamental shift that has occurred in statistics since 1980. The methods presented here were not possible before the birth of inexpensive computing in the 1980s. In addition, the automated collection of data has made large datasets comprised of highly accurate observations possible. GPS technology became

fully operational in 1994 and the deliberate distortion of GPS signals known as Select Availability ceased in 2000. High resolution habitat mapping through remote sensing is a product of the late-20th century. As the amount of data continue to increase and computing power becomes ever cheaper, computer intensive methods continue to take centre stage. Statistics researcher continue to produce more computer tools to aid in analysis.

Removing bias and measuring complex change accurately are of interest in many contexts. Computational Biology, Image Recognition, radio telemetry studies, and stock market volatility are just a few examples of methods being developed and becoming well known to small pockets of researchers. Looking at repurposing these methods is not trivial it involves learning the setting that the method was developed for and often reading research from several different disciplines in order to rewrite the code to work in the new context.

The methods outlined in this thesis required thousands of lines of code and a wide range of software packages. The high-end commercial packages MS Visual Studio (MS Visual Studio, 2005), MATLAB (MATLAB, 2012) and ArcGIS (ESRI, 2011) were used, as well as the free packages: Geographic Resources Analysis Support System (GRASS) (GRASS Development Team, 2012), R (R Core Team, 2012), Perl (Wall *et al.*, 2000), and WinBugs (Lunn *et al.*, 2000), and WinBugs (Lunn *et al.*, 2000). The iterative resampling method of bias correction the animal movement model was estimated in R and MATLAB and the results were displayed ArcGIS. The bias correction using BHM, the MCMC simulation was run in WinBugs, the observations were broken into spatial locations using ArcGIS and GRASS and the results were displayed using ArcGIS. The comparison using LDDMM the kernel smoothing was done with the R package ‘sm’ written primarily by Adrian Bowman (Bowman and Azzalini, 2010). The remaining computations and the graphs where done in MATLAB. Because the source code of the

implementation of LDDMM by the Bioinformatics Research Network is not open, an implementation of LDDMM was recreated in C++ and run from MATLAB using mex. The comparison using DD, the kernel smoothing was coded in MATLAB following Martinez and Martinez (2002) and the DD calculations were done in MATLAB using some functions from the image processing toolbox, some graphs were created in R. Throughout all of the work data were reformatted to move it between computer packages with Perl. The popular adage, "you don't know something until you teach it", can easily be modified to, "you don't know something until you code it."

The spatio-temporal analyses shown here cuts across the intersection of computer intensive statistics, biological statistics, econometrics and environmental statistics. The bias correction methods in Chapter 6 and 7 rely on frequentist and Bayesian frameworks respectively. Irrespective of the framework both utilise the long running computer simulations of computer intensive statistics and incorporate the biology of animal movement. The comparison of regression surfaces in Chapters 8 and 9 utilised computer intensive iterative estimates to determine the shortest smooth non-parametric curved paths connecting surfaces. They borrow from statistics developed for biological surfaces and apply them to environmental settings and econometrics settings. This intersection provides an area ripe for research because many methods can be borrowed from one field to another.

11 References

Bibliography

Aguilar, M. (2009). "A latent factor model of multivariate conditional heteroscedasticity." Journal of Financial Ecomometrics **7**(4): 481-503.

Ankudinova, J. and M. Ehrhardt (2008). "On the numerical solution of nonlinear Black–Scholes equations." Computers & Mathematics with Applications **56**(3): 799–812.

Arfken, G. (1985). The Method of Steepest Descents. Orlando, FL, Academic Press.

Ayache, E., P. Henrotte, S. Nassar and X. Wang (2004). "Can anyone solve the smile problem?" WILMOTT Magazine: 78--96.

Ayache, N. (2007). "Non-parametric Diffeomorphic Image Registration with the Demons Algorithm." Medical Image Computing and Computer-Assisted Intervention **4792**: 319-326.

Barnard, J., D. B. Rubin and N. Schenker (2005). Multiple imputation methods. Encyclopedia of Biostatistics. P. Armitage and T. Colton. West Sussex, England, Wiley. **5:** 3420-3428.

Beg, F., M. Miller, A. Trouve and L. Younes (2005). "Computing large deformation metric mappings via geodesic flows of diffeomorphisms. ." International Journal of Computer Vision **61**(2): 139–157.

Berliner, L. M. (1996). Hierarchical bayesian time series models. Maximum Entropy and Bayesian Methods. K. Hanson and R. Silver. Dordrecht, Kluwer Academic Publishers**:** 15-22.

Black, F. (1989). "How we came up with the option formula." The Journal of Portfolio Management **15**: 4-8.

Black, F. and M. Scholes (1973). "The pricing of options and corporate liabilities." Journal of Political Economy **81**(3): 637–654.

Blackwell, P. G. (1997). "Random diffusion models for animal movement." Ecological Modelling **100**(1-3): 87-102.

Bollerslev, T. (1986). "Generalized autoregressive conditional heteroskedasticity." Journal of Econometrics **31**(3): 307-327.

Bossard, M., J. Feranec and J. Otahel (2000). CORINE land cover technical guide - addendum 2000. E. E. A. T. Rep. **40**.

Bowman, A. W. (2007). "Comparing nonparametric surfaces." Statistical Modelling **6**: 1-21.

Bowman, A. W. and A. Azzalini (2010). R package '`sm`': nonparametric smoothing methods (version 2.2-4), University of Glasgow, UK and Università di Padova, Italia.

Boyce, W. E. and R. C. DiPrima (1997). Elementary Dierential Equations and Boundary Value Problems, Sixth Edition. Hoboken, NJ, Wiley.

Buckland, S. T. (1993). Distance Sampling: Estimating abundance of biological populations. London, Chapman & Hall.

Cachier, P., E. Bardinet, D. Dormont, X. Pennec and N. Ayache (2003). "Iconic feature based no rigid regristration: the PASHA algorithm." Computer Vision and Image Understanding **89**: 272–298.

Cachier, P., X. Pennec and N. Ayache (1999). "Fast non-rigid matching by gradient descent: study and improvement of the demons algorithm." Unité de recherche INRIA Sophia Antipolis **Theme 3**(3706).

Cacoullos, T. (1966). "Estimation of a multivariate density." Annals of the Institute of Statistical Mathematics **18**: 178-189.

Cain, I., J. W., P. R. Krausman, B. D. Jansen and J. R. Morgart (2005). "Influence of topography and GPS fix interval on GPS collar performance." Wildlife Society Bulletin **33**(3): 926-934.

Calvet, L. E. and A. J. Fisher (2004). "How to forecast long-run volatility: regime switching and the estimation of multifractal processes." Journal of Financial Econometrics **2**(1): 49—83.

Capinski, M. and T. Zastawniak (2003). Mathematics for Finance: An introduction to financial engineering. London, Springer-Verlag.

Castro, F. S., C. Pollo, O. Cuisenaire, J. Villemure and J. Thiran (2006). "Validation of experts versus atlas-based and automatic registration methods for subthalamic nucleus targeting on MRI." International Journal of Computer Assisted Radiology and Surgery **1**(1): 5-12.

Chicago Board Options Exchange, Incorporated (2012) "S&P 500® (SPX SM) INDEX OPTIONS." **March 2012**. Accessed: 26/04/2012, http://www.cboe.com/micro/spx/pdf/SPX_QRG2.pdf.

Cleveland, W., E. Grosse and W. Shyu (1992). Local regression models. Statistical Models in S. J. Chambers and T. Hastie. Pacific Grove, California, Wadsworth and Brooks/Cole. **309–76**.

Codling, E. A., M. J. Plank and S. Benhamou (2008). "Random walk models in biology." Journal of the Royal Society Interface **5**: 813–834.

Cont, R. and J. Da Fonsecca (2001). Deformation of Implied Volatility Surfaces: An empirical analysis. Tokyo, Springer.

Cont, R. and J. da Fonsecca (2002). "Dynamics of implied volatility surfaces." Quantitative Finance **2**: 45-60.

Cont, R., J. Da Fonsecca and V. Durrleman (2002). "Stochastic models of implied volatility surfaces." Economic Notes **31**(2): 361-377.

Cressie, N. (1990). "The origins of kriging." Mathematical Geology **22**: 239–252.

Cressie, N. and C. K. Wikle (2011). Spatial-temporal Data. Hoboken, NJ, John Wiley and Sons, Inc.

D'Eon, R. G., R. Serrouya, G. Smith and C. O. Kochanny (2002). "GPS radiotelemetry error and bias in mountainous terrain." Wildlife Society Bulletin **30**(2): 430-439.

Deheuvel, P. (1977). "Estimation non paramétrique de la densité par histogrammes généralisés (II)." Revue de Statistique Appliquée **25**(3): 5-42.

Delbaen, F. and W. Schachermayer (1994). "A general version of the fundemental theorem of asset pricing." Mathematische Annalen **300**(1): 463–520.

Derman, E. (1999). "Regimes of volatility." RISK **April**.

Dette, H. and A. Munk (1998). "Nonparametric comparison of several regression functions: exact and asymptotic theory." The Annals of Statistics **26**(6): 2339-2368.

Dette, H. and N. Neumeyer (2003). "Nonparametric comparison of regression curves: an empirical process approach." The Annals of Statistics **31**(3): 880-920.

Devroye, L. (1983). "The equivalence of weak, strong, and complete convergence in L_1 for kernel density estimates." The Annals of Statistics **11**: 896-904.

Devroye, L. and C. S. Penrod (1984). "The consistency of automatic kernel density estimates." The Annals of Statistics **12**: 1231-1249.

Doz, C. and E. Renault (2006). "Factor stochastic volatility in mean models: a GMM approach." Econometric Reviews **25**: 275-309.

DuBois, J. L., R. P. Multhauf and C. A. Ziegler (2002). The Invention and Development of the Radiosonde. Washington, DC, Smithsonian Institution Press, from http://hdl.handle.net/10088/2453.

Dupuis, P., U. Grenander and M. I. Miller (1998). "Variational problems on flows of diffeomorphisms for image matching." Quarterly of Applied Mathematics **LVI**: 587–600.

Edenius, L. (1997). "Field test of a GPS location system for moose *Alces alces* under Scandinavian boreal conditions." Wildlife Biology **3**: 39-43.

Efron, B. and R. Tibshirani (1991). "Statistical data analysis in the computer age." Science **253**(5018): 390-395.

Einstein, A. (1956). Investigations on the Theory of the Brownian Movement. New York, NY, Dover.

Engle, R. F. (1982). "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation." Econometrica **50**: 987-1008.

Epanechnikov, V. A. (1969). "Non-parametric estimation of a multivariate probability density."

Theory of Probability and its Applications **14**: 153–158.

ESRI (2011). ArcGIS Desktop: Release 10. Redlands, CA, Environmental Systems Research Institute, http://www.esri.com/software/arcgis.

Fisher, N. I. (1993). Statistical Analysis of Circular Data. Cambridge. England, Cambridge University Press.

Frair, J. L., J. Fieberg, M. Hebblewhite, F. Cagnacci, N. J. DeCesare and L. Pedrotti (2010). "Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data." Philosophical Transactions of the Royal Society B: Biological Sciences **365**(1550): 2187-2200.

Frair, J. L., S. E. Nielsen, E. H. Merrill, S. R. Lele, M. S. Boyce, R. H. M. Munro, et al. (2004). "Removing GPS collar bias in habitat selection studies." Journal of Applied Ecology **41**: 201-212.

Fryer, M. J. (1976). "Some errors associated with the nonparametric estimation of density functions." Journal of the Institute of Mathematics and Its Applications **18**: 371-380.

Gallian, J. A. (2010). Contemporary abstract algebra. Belmont, CA, Brooks/Cole, Cengage Learning.

Gau, R. J., R. Mulders, L. M. Ciarniello, D. C. Heard and et al. (2004). "Uncontrolled field performance of Televilt GPS-SimplexTM collars on grizzly bears in western and northern Canada." Wildlife Society Bulletin **32**(3): 693-701.

Gentle, J. E., W. Härdle and Y. Mori (2004). Handbook of Computational Statistics. Berlin, Springer-Verlag.

GRASS Development Team (2012). Geographic Resources Analysis Support System (GRASS) Software, Version 6.4.1., Open Source Geospatial Foundation., http://grass.osgeo.org.

Graves, T. A. and J. S. Waller (2006). "Understanding the causes of missed global positioning system telemetry fixes." The Journal of Wildlife Management **70**(3): 844-851.

Grenander, U. and M. I. Miller (1998). "Computational anatomy: An emerging discipline." Quarterly of Applied Mathematics **56**(4): 617-694.

Grünbaum, D. (1999). "Advection–diffusion equations for generalized tactic searching behaviors." Journal of Mathematical Biology **38**(2): 169-194.

Hall, P. (1989). "On convergence rates in nonparametric problems." International Statistical Review **57**: 45-58.

Hall, P. and E. J. Hannan (1988). "On stochastic complexity and nonparametric density estimation." Biometrika **75**: 705-714.

Hansen, P. R. and A. Lunde (2005). "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?" Journal of Applied Econometrics **20**(7): 873-889.

Hardy, M. R. (2001). "A regime-switching model for long-term stock returns." North American Actuarial Journal **5**(2): 41-53.

Harford, T. (2012) "Black-Scholes: The maths formula linked to the financial crash." 27 April 2012. Accessed: 29 April 2012, http://www.bbc.co.uk/news/magazine-17866646.

Hastie, T. and R. Tibshirani (1990). Generalized Additive Models. London S. Chapman & Hall.

Hernandez, M., S. Olmos and X. Pennec (2008). "Comparing algorithms for diffeomorphic registration: Stationary LDDMM and Diffeomorphic Demons." 2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy: 24-35.

Holtz, R. L. (2011) "Atlas gives scientists new view of the brain." Wall Street Journal April 13. Accessed: 1 December 2011, http://online.wsj.com/article/SB10001424052748703518704576258842075419366.html.

Horne, J. S., E. O. Garton and K. A. Sager-Fradkin (2007). "Correcting home-range models for observation bias." Journal of Wildlife Management **71**(3): 996-1001.

Hui, G. and H. I. AlDarmaki (2012). Editing and imputation of the 2011 Abu Dhabi census. Conference of European Statisticians - Work Session on Statistical Data Editing. Oslo, Norway, United Nations Economic Commission for Europe.

Hull, J. C. (2011). Options, Futures and Other Derivatives (8th Edition). Upper Saddle River, New Jersey, Prentice Hall

Hurvich, C. M., J. S. Siminoff and C.-L. Tsai (1998). "Smoothing parameter selection is nonparametric regression using as improved Akaike information criterion." Journal of Royal Statistical Society, Series B **60**: 271-293.

Hurvich, C. M. and C. L. Tsai (1989). "Regression and time series model selection in small samples." Biometrika **76**: 297-307.

Izenman, A. J. (1991). "Recent developments in nonparametric density estimation." Journal of the American Statistical association **86**(413): 205-224.

Johnson, B. K., A. A. Ager, S. L. Findholt, M. J. Wisdom, D. B. Marx, J. W. Kern, et al. (1998). "Mitigating spatial differences in observation rate of automated telemetry systems." The Journal of Wildlife Management **62**(3): 958-967.

Jonsen, I. D., R. A. Myers and J. M. Flemming (2003). "Meta-analysis of animal movement using state-space models." Ecology **84**(11): 3055-3063.

Joshi, S. C., J. Wang, M. I. Miller, D. V. Essen and U. Grenander (1995). "On the differential geometry of the cortical surface." Proceedings of SPIE 1995 International Symposium: Vision Geometry IV, San Diego, CA: 304--311.

Kareiva, P. M. and N. Shigesada (1983). "Analyzing insect movement as a correlated random walk." Oecologia **56**(2/3): 234-238.

Kendall, D. G. (1984). "Shape manifolds. Procrustean metrics and complex projective spaces." Bulletin London Math. Soc. **16**(2): 81-121.

Krige, D. G. (1951). A statistical approach to some mine valuations and allied problems at the Witwatersrand, University of Witwatersrand. **Masters**.

Kroon, D.-J. and C. H. Slump (2009). MRI modalitiy transformation in demon registration. International Symposium on Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE Boston, MA.

Kufeld, R. C., D. C. Bowden and J. M. Siperek (1987). "Evaluation of a telemetry system for measuring habitat usage in mountainous terrain." Northwest Science **61**: 249–256.

Leff, H. S. and A. F. Rex (1990). Maxwell's Demon: Entropy, Information, Computing. Bristol, Adam-Hilger.

Lewis, J. S., J. L. Rachlow, E. O. Garton and L. A. Vierling (2007). "Effects of habitat on GPS collar performance: Using data screening to reduce location error." Journal of Applied Ecology **44**(3): 663-671.

Lunn, D. J., A. Thomas, N. Best and D. Spiegelhalter (2000). "WinBUGS -- a bayesian modelling framework: concepts, structure, and extensibility." Statistics and Computing **10**: 325--337.

MacKenzie, D. I. (2006). "Modeling the probability of resource use: the effect of, and dealing with, detecting a species imperfectly." Journal of Wildlife Management **70**: 367-374.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle and C. A. Langtimm (2002). "Estimating site occupancy rates when detection probabilities are less than one." Ecology **83**: 2248-2255.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, J. E. Hines and L. L. Bailey (2006). Occupancy Estimation and Modeling: Inferring patterns and dynamics of species occurrence. . San Diego, USA, Elsevier.

MacKenzie, D. I., J. D. Nichols, M. E. Seamans and R. J. Gutierrez (2009). "Modeling species occurrence dynamics with multiple states and imperfect detection." Ecology **90**: 823-835.

MacKenzie, D. I., J. D. Nichols, N. Sutton, K. Kawanishi and L. L. Bailey (2005). "Improving inferences in population studies of rare species that are detected imperfectly." Ecology **86**: 1101-1113.

Manly, B. F. J., L. L. McDonald, D. L. Thomas, T. L. McDonald and W. P. Erickson (2002). Resource Selection by Animals: Statistical Design and Analysis for Field Studies. Dordrecht, The Netherlands, Kluwer Academic Publishers.

Marron, J. S. and D. Nolan (1987). Canonical kernels for density estimation. Technical Report. Chapel Hill, University of North Carolina.

Marsh, L. (1988). "The form and consequences of random walk movement models." Journal of Theoretical Biology **133**(1): 113.

Martinez, W. L. and A. R. Martinez (2002). Computational Statistics Handbook with MATLAB. Boca Raton, Chapman & Hall / CRC.

MATLAB (2012). R2012a (7.14.0). Natick, Massachusetts, The MathWorks Inc, http://www.mathworks.com/products/matlab/.

McGraw-Hill Concise Encyclopedia of Science and Technology (2005). Maxwell's Demon. G. V. R. Library. New York, McGraw-Hill Professional**:** 1340.

Merton, R. C. (1973). "Theory of rational option pricing." Bell Journal of Economics and Management Science **4**: 141-183.

Miller, M. I., A. Trouve and L. Younes (2002). "On the metrics and Euler-Lagrange equations of

computational anatomy." Annual Review of Biomedical Engineering, **4**(375-405).

Moen, R., J. Pastor, Y. Cohen and C. C. Schwartz (1996). "Effects of moose movement and habitat use on GPS collar performance." The Journal of Wildlife Management **60**: 569-668.

Morales, J. M., D. T. Haydon, J. Frair, K. E. Holsinger and J. M. Fryxell (2004). "Extracting more out of relocation data: Building movement models as mixtures of random walks." Ecology **85**(9): 2436-2445.

MS Visual Studio (2005). Visual Studio 2005 (9.0). Redmond, Washington, Microsoft Corporation, http://www.microsoft.com/visualstudio/.

Nams, V. O. (1989). "Effects of radiotelemetry error on sample size and bias when testing for habitat selection." Canadian Journal of Zoology **67**(7): 1631-1636.

NBS (2011). Censuses, United Arab Emirates National Bureau Of Statistics, from http://www.uaestatistics.gov.ae/CensusEN/tabid/202/Default.aspx.

Newton, H. (2011). Newton's Telecom Dictionary. New York, Flatiron Publishing.

Nielson, R. M., B. F. J. Manly, L. L. McDonald, H. Sawyer and T. L. McDonald (2009). "Estimating habitat selection when GPS fix success is less than 100%." Ecology **90**(10): 2956-2962.

Nussbaum, M. (1996). "Asymptotic equivalence of density estimation and gaussian white noise." The Annals of Statistics **24**(6): 2399 – 2430.

Parzen, E. (1962). "On estimation of a probability density function and mode." The Annals of Mathematical Statistics **33**: 1065-1076.

Patterson, T., L. Thomas, C. Wilcox, O. Ovaskainen and J. Matthiopoulos (2008). "State-space models of individual animal movement." Trends Ecol Evol. **23**(2): 87-94.

Pennec, X., P. Cachier and N. Ayache (1999). Understanding the "Demon's Algorithm": 3D non-rigid negistration by gradient descent. In 2nd International Conference on Medical Image Computing and Computer-assisted Intervention (MICCAI'99) LNSC, Cambridge, UK.

Psychoyios, D., G. Dotsis and R. N. Markellos (2010). "A jump diffusion model for VIX volatility options and futures." Review of Quantitative Finance and Accounting **35**(3): 245-269.

R Core Team (2012). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, http://www.R-project.org.

Rempel, R. S., A. R. Rodgers and K. F. Abraham (1995). "Performance of a GPS animal location system under boreal forest canopy." The Journal of Wildlife Management **59**(3): 543-551.

Rettie, W. J. and P. D. McLoughlin (1999). "Overcoming radiotelemetry bias in habitat-selection studies." Journal of Zoology **77**: 1175-1184.

Robb, K. W. and D. F. Mayers (1996). Biomedical Imaging, Visualization, and Analysis. New York, NY, John Wiley and Sons, Inc.

Robert, A. (1981). "A stable numerical integration scheme for the primitive meteorological equations." Atmosphere-Ocean **19**: 35-46.

Rosenblatt (1956). "Remarks on some nonparametric estimates of a density function." Annals of Mathematical Statistics **27**: 832-837.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York, Wiley.

Rumble, M. A. and F. Lindzey (1997). "Effects of forest vegetation and topography on global positioning system collars for elk." Resource Technology Institute Symposium **4**: 492 - 501.

Sauer, J. R., B. G. Peterjohn and W. A. Link (1994). "Observer differences in the North American breeding bird survey." The Auk **111**(1): 50-62.

SCAD (2011). Abu Dhabi in figures 2011. Abu Dhabi, Statistics Centre - Abu Dhabi, 4/7/2012, from http://www.scad.ae/SCAD%20Publications/Abu%20Dhabi%20in%20Figures%202011%20Eng.pdf.

Scott, D. W. (1992). Multivariate Density Estimation: Theory, Practice, and Visualization, Wiley, New York.

Seidelmann, P. K., Ed. (1992). Explanatory Supplement to the Astronomical Almanac. Sausalito, CA, University Science Books.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. New York, Chapman and Hall.

Simpson, E. H. (1951). "The interpretation of interaction in contingency tables." Journal of the Royal Statistical Society, Ser. B **13**: 238-241.

Standard & Poor's Financial Services LLC (2012) "S&P indices." Accessed: 26/4/2012, http://www.standardandpoors.com/indices/articles/en/us/?articleType=PDF&assetID=1221190434733.

Staniforth, A. and J. Cote (1991). "Semi-Lagrangian integration schema for atmospheric models – a review." Monthly Weather Review **119**: 2206-2223.

The Royal Swedish Academy of Sciences. (1997). "The Prize in Economics 1997 - Press Release." Retrieved 29 Apr, 2012, from http://www.nobelprize.org/nobel_prizes/economics/laureates/1997/press.html.

Thirion, J. P. (1994). Extremal points: definition and application to 3D image registration. IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA.

Thirion, J. P. (1995). Fast non-rigid matching of 3D medical images. Medical Robotics and Computer Aided Surgery (MRCAS '95), Baltimore, MD.

Thirion, J. P. (1998). "Image matching as a diffusion process: An analogy with Maxwells's demons." Medical Imaging Analysis **2**(3): 243-260.

Thurfjell, H., J. P. Ball, P.-a. Åhlén, P. Kornacher, H. Dettki and K. Sjöberg (2009). "Habitat use and spatial patterns of wild boar *Sus scrofa* (L.): Agricultural fields and edges." European Journal of Wildlife Research **55**(5): 517-523.

Trouvé, A. (1998). "Diffeomorphisms groups and pattern matching in image analysis." International Journal of Computer Vision **28**(3): 213-221.

Vercauteren, T., X. Pennec, A. Perchant and N. Ayache (2007). "Non-Parametric diffeomorphic image registration with the demons algorithm." Med Image Comput Comput Assist Interv **10**(Pt 2): 319-326.

Vercauteren, T., X. Pennec, A. Perchant and N. Ayache (2009). "Diffeomorphic demons: Efficient non-parametric image registration." NeuroImage **45**(1, Supplement 1): S61-S72.

Villepique, J. T., V. C. Bleich, B. M. Pierce, T. R. Stephenson, R. A. Botta and R. T. Bowyer (2008). "Evaluating GPS collar error: A critical evaluation of Televilt POSREC-Science™ collars and a method for screening location data." California Fish Game **94**: 155–168.

Wall, L., T. Christiansen and J. Orwant (2000). Programming Perl, 3rd Edition. Sebastopol, CA, O'Reilly Media.

Wand, M. P. and M. C. Jones (1995). Kernel Smoothing. London Chapman and Hall/CRC.

Wang, H., L. Dong, J. O'Daniel, R. Mohan, A. S. Garden, K. K. Ang, et al. (2005). "Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy." Physics in Medicine and Biology **50**(12): 2887 - 2905.

Wasserman, L. (2006). All of Nonparametric Statistics. New York, Springer Science + Business Media, Inc.

Wegman, E. J. (1988). "Computational statistics: A new agenda for statistical theory and practice." Journal of the Washington Academy of Sciences **78**: 310-322.

Welander, P. (1955). "Studies on the general development of motion in a two-dimensional, ideal fluid." Tellus **7**(2): 141-156.

Whaley, R. E. (1993). "Derivatives on market volatility: Hedging tools long overdue." Journal of Derivatives 1 **Fall 1993**: 71–84.

Whaley, R. E. (2009). "Understanding the VIX." The Journal of Portfolio Management **35**(3): 98-105.

White, G. C. and R. A. Garrott (1990). Analysis of Wildlife Radio-Tracking Data. San Diego, Academic Press.

Wu, H.-i., B.-L. Li, T. A. Springer and W. H. Neill (2000). "Modelling animal movement as a persistent random walk in two dimensions: Expected magnitude of net displacement " Ecological Modelling **132**(1–2): 115-124.

Zwillinger, D. and S. Kokoska (2000). Standard Probability and Statistics Tables and Formulae. Boca Raton, Champan and Hall / CRC.